MECHANISMS OF BACTERIAL POLYPRENYL TRANSFERASES AND MULTI-TARGET DRUG DISCOVERY FOR TUBERCULOSIS

BY

XINXIN FENG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Eric Oldfield, Chair
Professor Yi Lu
Professor Robert B. Gennis
Professor Satish K. Nair

Abstract

Bacterial polyprenyl transferases transfer polyprenyl groups onto molecular acceptors, such as proteins and small molecules, and they are very important functional entities that are involved in bacterial cell wall biosynthesis, biofilm formation, virulence formation and natural product formation. Structurally, bacterial polyprenyl transferases can be categorized into eight groups: $\alpha_{cyclase}$, α, αβ, αβγ, βγ, ε, ζ and TIM barrel fold proteins. Six bacterial polyprenyl transferases were investigated in the current research, including *Streptomyces ghanaensis* MoeO5 and *Bacillus subtilis*/*Staphylococcus aureus* PcrB (TIM barrel fold), *Bacillus subtilis* YisP (ε fold), *Bradyrhizobium japonicum* Kaurene synthase ($\alpha_{cyclase}$ fold), *Mycobacterium tuberculosis* Rv3378c and DPPS (ζ fold The function, mechanism of action, structure, inhibition and functional engineering of these six polyprenyl transferases were studied by X-ray crystallography, mutagenesis, activity assays, thermal dynamics measurements *etc*. Besides that, a multi-target drug discovery approach was also proposed as an attempt to combat drug resistance in anti-infective drug discovery campaigns, both based on structural homology of *Mycobacterium tuberculosis* Rv3378c and DPPS (ζ fold), and the multi-target effect of SQ-109 and its analogs.

MoeO5 catalyzes the first step biosynthesis of Moenomycin A, which is a phosphoglycolipid antibiotic. We determined the first crystal structure of MoeO5 and studied the binding mode of bound substrates. We also assayed its activity with different substrates and compared MoeO5 with its homolog PcrB, which led to a proposal of MoeO5 reaction mechanism.

YisP is essential in biofilm formation in *Bacillus subtilis* and is predicted to produce C_{30} isoprenoids. We determined the structure of YisP, and proved that YisP acts as a phosphatase. Using DSC, we confirmed that farnesol shift and broaden the gel-to-liquid crystal transition of lipids.

The X-ray structure of the bacterial diterpene cyclase *ent*-kaur-16-ene synthase from the soil bacterium *Bradyrhizobium japonicum* was determined in apo-, substrate and inhibitor-bound forms. The catalytic activity of the cyclase was studied by site-directed mutagenesis and inhibition of the enzyme was also investigated.

Structures of tuberculosinol/*iso*-tuberculosinol synthase (Rv3378c) and *cis*-decaprenyl diphosphate synthase (DPPS) from *M. tuberculosis* were determined. They are targets for anti-infective therapies that block virulence factor formation and for cell wall biosynthesis, respectively. Given the similarity in local and global structure between these two proteins, the possibility exists that it may be possible to develop inhibitors that target not only virulence, but also cell wall biosynthesis.

SQ-109, one new anti-tuberculosis ethylenediamine drug, was found to be active against other bacteria and yeasts. Based on the SAR study and biochemistry assay results with eleven SQ-109 analogs, we found that it might be an example of multi-target drug candidate that inhibits isoprenoid/quinone biosynthesis and/or respiration.

To My Parents and My Husband

Acknowledgments

It is my pleasure to express my gratitude to people who have been helping me throughout my Ph.D. studies.

First of all, I would like to thank my research advisor, Prof. Eric Oldfield for his constant support and guidance. He is a great mentor who teaches me how to become a good researcher. I am really grateful for the opportunity he gave me to work in the lab and the trust he has in me to pursue my own research ideas. It has been a really great experience interacting with Dr. Oldfield, both scientifically and personally.

Additionally, I am grateful to the members of the Oldfield group. I would like to thank Dr. Yi-liang Liu, Dr. Wei Zhu, Dr. Ke wang, Dr. Rong Cao, Dr. Fu-Yang Lin, Dr. Weixue Wang, Dr. Kai Li and Dr. Yonghui Zhang for training me in the lab, answering all my questions, and assisting me with my experiments. I also thank Francisco and Jikun for maintaining our internal website and make data processing much easier. I thank Janish, Yang and Guodong for encouraging me along my way, and my undergrad student Shannon for working with me on various projects.

Moreover, I am very grateful to my collaborators from University of Illinois and other campus: Prof. Rey-Ting Guo, Dr. Tzu-Ping Ko, Prof. Michael Cynamon, Prof. Douglas Mitchell, Prof. Robert Gennis and Prof. Dean Crick, for their help and input on X-ray crystallography, cell-based assays, respiration assays and molecular dynamics simulation.

Furthermore, I am very thankful to my thesis committee members, Prof. Robert Gennis, Prof. Yi Lu, and Prof. Satish K. Nair for their suggestions during the committee meetings, as well as advice and support for my current and future career.

Last but not least, I would like to thank my parents, Jingquan Feng and Guijian Zheng for their support and unconditional love throughout these years. I would also thank all my dear friends for years of friendship, encouragement and support. And lastly, I sincerely thank my husband, Yugang Bai, for his support and company on my entire journey.

Table of Contents

Chapter 1: Introduction to Bacterial Polyprenyl Transferases and Drug Discovery for Tuberculosis

1.1 Bacterial Polyprenyl Transferases

Polyprenyl transferases refer to proteins that transfer polyprenyl groups from prenyl donors to prenyl acceptors. As shown in Figure 1.1, the prenyl donors are usually polyisoprenyl diphosphates that donate their allylic polyisoprenyl groups to prenyl acceptors, such as other proteins (protein prenylation), small molecules (small molecule prenylation), polyisoprenyl diphosphates (polyprenyl diphosphate synthesis) or themselves (cyclization).

Protein prenylation is a type of post-translational modification of proteins, in which farnesyl (15-carbon) or more commonly geranylgeranyl (20-carbon) isoprenoids are added via thioether linkages to cysteine residues at or near the C terminus of intracellular proteins.[1] Protein prenylation is an important process to mediate protein-protein interactions and protein-membrane interactions, as the attached lipid is required for proper function of the modified protein, either as a mediator of membrane association or a determinant for specific protein-protein interactions. Three distinct protein prenyltransferases can be classified in two functional classes[1]: the C*AAX* prenyltransferases, identified by their lipid substrate and termed protein farnesyltransferase (FTase) and protein geranylgeranyltransferase type I (GGTase-I); and the Rab geranylgeranyltransferase or protein geranylgeranyltransferase type II (GGTase-II).

Small molecule prenylation is very common in lipid formation and natural product formation in all organisms. For example, *Archaeoglobus fulgidus* geranylgeranylglyceryl phosphate synthase (AfGGGPS)[2] catalyzes the reaction between geranylgeranyl diphosphate (GGPP) and glycerol-1-phosphate (G1P) to form geranylgeranylglyceryl phosphate (GGGP), a precursor of the (perhydro)geranylgeranyl-based phospholipids found in Archaea[3]. Also, In *Streptomyces ghanaensis*, the enzyme MoeO5, encoded by the *moe* gene cluster 1 catalyzes the initial step of moenomycin production in which the C_{15}-hydrocarbon moiety of farnesyl pyrophosphate (FPP) is transferred to the 2-hydroxyl group of 3-phosphoglycerate (3PG), forming an ether bond. The phosphoglycolipid

antibiotic moenomycin directly blocks bacterial cell wall biosynthesis by inhibiting peptidoglycan glycosytransferases.[4]

Polyprenyl diphosphate synthesis and cyclization polyprenyl diphosphate belong to a broader class of reactions termed terpene biosynthesis. Terpenes (or terpenoids) represent the largest class of small molecules on the planet[5] and most are synthesized by enzymes that contain one or more of six main types of fold: α, β, γ, δ, ε and ζ[6] (Figure 1.2). Functionally, they can be classified into the following groups: head-to-tail *trans* prenyl transferases (αα- and αδ-domain structures), head-to-tail *cis* prenyl transferases (ζ-domain structures), head-to-head prenyl transferases (ε-domain structures), diterpene cyclases (αβγ-fold structures).

Bacterial polyprenyl transferases are very important functional entities that are involved in bacterial cell wall biosynthesis, biofilm formation, virulence formation and natural product formation.[6-7] Six bacterial polyprenyl transferases were investigated in the current research (Scheme 1.1), including *Streptomyces ghanaensis* MoeO5[8] and *Bacillus subtilis*/*Staphylococcus aureus* PcrB[9] (TIM barrel fold), *Bacillus subtilis* YisP[10] (ε fold), *Bradyrhizobium japonicum* Kaurene synthase[11] ($\alpha_{cyclase}$ fold), *Mycobacterium tuberculosis* Rv3378c and DPPS[12] (ζ fold) (Figure 1.3), on their functions, mechanisms of action, structures, inhibition and functional engineering studied by X-ray crystallography, mutagenesis, activity assay, thermal dynamic measurement *etc*.

1.2 *Streptomyces ghanaensis MoeO5 and Bacillus subtilis/Staphylococcus aureus PcrB*

Recently, several new structures of PTs that contain a TIM barrel fold (geranylgeranylglyceryl phosphate synthase (GGGPS), MoeO5 and PcrB) have been discovered[2, 8, 13], and there is considerable interest in their mechanisms of action[8, 13b]. The reactions catalyzed by these three proteins are shown in Scheme 1.2.

MoeO5 catalyzes the initial step of moenomycin production in which the C_{15}-hydrocarbon moiety of farnesyl pyrophosphate (FPP) is transferred to the 2-hydroxyl group of 3-phosphoglycerate (3PG), forming an ether bond. As showed in Scheme 1.2, the reaction is similar to that of GGGPS for synthesizing Archaea-type phospholipids.[2] However, unlike GGGPS whose product retains the all-*trans* configuration of the

isoprenyl chain, the farnesyl group transferred by MoeO5 undergoes a *trans*-to-*cis* isomerization at the C2-C3 double bond.[13b]

Then enzyme PcrB catalyzes a very similar reaction as that of GGGPS, but in this case the prenyl donor is heptaprenyl diphosphate (HepPP)[14]. The PcrB product, heptaprenylglyceryl phosphate (HepGP), is then dephosphorylated to heptaprenylglycerol, then acetylated, Scheme 1.1[14]. The function of this ester product is unknown. However, PcrB forms part of an operon with PcrA, an essential DNA helicase[15], and a Δ*pcrB* mutant (in *Bacillus subtilis*) has a deranged, "cloggy" pattern of growth[14], so it seems likely that PcrB is important for growth (or perhaps, virulence) in other Bacilli, such as the pathogen *Staphylococcus aureus*, and may be of interest as a drug target.

The crystal structures of GGGPS and the bacterial homologue PcrB reveal a triose phosphate isomerase (TIM)-barrel fold, which had not been observed previously in prenyltransferases (PTs).[2, 13a, 14] The sequences of GGGPS and PcrB share 35% amino acid identity, but MoeO5 shows only 10% identity to both enzymes. We thus investigated the structures and mechanisms of action of MoeO5 and PcrB, in comparison of those of GGGPS.

1.3 *Bacillus subtilis* YisP

There is currently considerable interest in the structures, functions and inhibition of the enzymes catalyze the "head-to-head" condensation of isoprenoid diphosphates to form tri- terpenes such as squalene and dehydrosqualene, such as squalene synthase (SQS) and dehydrosqualene synthase (CrtM) (Scheme 1.3a). Squalene synthase inhibition represents new mechanisms of action to develop cholesterol-lowering drugs[16] and antiparasitics targeting ergosterol biosynthesis[17]. CrtM from *Staphylococcus aureus* is involved in formation of staphyloxanthin[18], a virulence factor for the organism providing resistance to killing from host reactive oxygen species.[19] Thus, CrtM is of interest as a target in anti-virulence therapy.

The CrtM gene (or a close analog) is also present in *Bacillus subtilis* (*yisP*) and it was recently reported[20] that YisP was involved in biofilm and pigment formation and had squalene synthase activity which was blocked by zaragozic acid (Scheme 1.3b), a known SQS/CrtM inhibitor.[21] Zaragozic acid was also reported to block biofilm formation in *S.*

aureus, but had no effect on cell growth in either organism. Moreover, one of the two conserved active site motifs in SQS/CrtM family is missing in YisP. We thus investigated the structure, function and mechanism of action of YisP with X-ray crystallography and activity assay.

1.4 *Bradyrhizobium japonicum* Kaurene synthase

Diterpenes have been found in bacteria[22], and these compounds are of interest since some have activity as anti-infective and anti-cancer drug leads, virulence factors, as well as plant growth hormones. There are two different classes of terpene cyclases, Class I and Class II, defined by the amino acid motifs that are essential for catalysis[23]. Class I cyclases contain two conserved motifs, **D**DXX**D**, and (**N,D**)DXX(**S,T,G**)XXX(**E,D)** (hereafter referred to as NTE), in which the boldface residues generally bind to the three Mg^{2+} that facilitate ionization of an isoprenoid diphosphate group, generating a reactive carbocation intermediate. In contrast, Class II cyclases typically contain only a DX**D**D motif in which the boldface aspartic acid protonates an isoprenoid double bond or an oxirane to, once again, generate a reactive carbocation intermediate. In some cases, both Class I as well as Class II motifs are present and these molecules are bi-functional or in one reported case, tri-functional[24]. For example, abietadiene synthase from *Abies grandis* (AgABS) is bi-functional and converts geranylgeranyl diphosphate (GGPP) to copalyl diphosphate (CPP) in a Class II reaction and then CPP is converted to abietadiene in a Class I reaction, Scheme 1.4a. Similarly, the moss *Physcomitrella patens* utilizes a trifunctional CPPS/KS (PpCPPS/KS) to convert GGPP to *ent*-CPP in a Class II reaction and then to *ent*-kaur-16-ene (20%) and 16α-hydroxy-*ent*-kaurane (80%) via a Class I reaction with, in the case of kauranol, carbocation quenching[25], Scheme 1.4b.

Here, we report the first structure of a bacterial diterpene cyclase, *ent*-kaurene synthase from the soil bacterium *Bradyrhizobium japonicum* (BjKS; gene # blr2150)[22b]. BjKS catalyzes the cyclization of *ent*-CPP which is produced by *ent*-copalyl diphosphate synthase (*ent*-CPPS; blr2149) from GGPP (Scheme 1.4c), and one possible mechanism of action [26] for the *ent*-kaurene synthase reactions is shown in Scheme 1.4c. The *ent*-KS reactions proceed via initial diphosphate loss from *ent*-CPP (catalyzed by Mg^{2+}) and cyclization to form the 8-carbonium ion, the pimaren-8-yl cation, which then undergoes

secondary cyclization with Wagner-Meerwein migration of C-12 to C-16 to form the *ent*-kauran-16-yl cation. This cation can then loose a proton to form *ent*-kaur-16-ene, or in *P. patens* the *ent*-kauranyl cation can also be quenched by a water molecule to form 16-α-hydroxy-*ent*-kaurane, Scheme 1.4b[25].

Structurally, Class I proteins contain a catalytic α domain while Class II proteins contain two domains, βγ, and with bi-functional species such as abietadiene synthase, all three domains (αβγ) are present[27]. In previous work[7b] it was proposed that these αβγ domain proteins might have arisen by the fusion of the genes of α+βγ proteins in ancestral, perhaps soil dwelling, bacteria, so determining the structures of bacterial α (and βγ) cyclases is of interest not only from a mechanism of action perspective, but also from an evolutionary one. Is, for example, the structure of BjKS (a predicted[7b] α domain protein) most similar to that of the α domains seen in modern plant terpene cyclases? To investigate this and other questions we determined the structure of BjKS in the presence of its *ent*-CPP substrate, as well as apo- and inhibitor-bound structures, and we used site-directed mutagenesis to probe its catalytic activity. In addition, we compared the BjKS structure with the structures of diverse other terpene cyclases, and synthases.

1.5 *Mycobacterium tuberculosis* Rv3378c and decaprenyl diphosphate synthase

Tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, is a major source of morbidity and mortality worldwide with almost two million deaths annually, and the rise in multi-drug resistant and extensively drug resistant strains is of great concern[28]. There is thus considerable interest in the development of new drugs and new leads such as SQ109 and TMC-207[29], which target cell wall biosynthesis and ATP formation, respectively, and also in the development of new therapeutic approaches that target virulence factor (VF) formation. VFs are by definition not essential for bacterial growth outside of host cells, but are involved in processes such as invasion, persistence, lysis, and evasion of innate immune system responses, as found for example with staphyloxanthin in *Staphylococcus aureus*[30]. This carotenoid VF provides resistance to host-based reactive oxygen species-based killing, and inhibiting staphyloxanthin biosynthesis is a novel route to anti-infective therapy[30b]. In *M. tuberculosis*, one class of VFs are the tuberculosinols[31]: tuberculosinol (TOH) and the (±) *iso*-tuberculosinols (*iso*-

TOH), Scheme 1.5. The latter were first isolated from a marine sponge from the Nosy Be island, Madagascar, and are thus also known as (±) nosyberkols[32] (Scheme 1.5). The original structures proposed[31-32] have been confirmed by total synthesis[33]. The biosynthesis of the tuberculosinols is catalyzed by two enzymes: Rv3377c, tuberculosinyl (halama-5,13-dien-15-yl) diphosphate synthase, and Rv3378c, tuberculosinol/(±)*iso*-tuberculosinol synthase (Scheme 1.5). Both proteins are essential for bacterial survival inside macrophages[34] with the tuberculosinols inhibiting phagolysosome maturation, as well as macrophage phagocytosis[31b, c, 35]. The three-dimensional structures of both Rv3377c and Rv3378c are thus of considerable interest since they are thought to represent novel anti-virulence therapeutic targets for tuberculosis[36]. Rv3377c is a diterpene cyclase that converts geranylgeranyl diphosphate (GGPP) into tuberculosinyl diphosphate (TPP), while Rv3378c converts TPP into TOH and the *iso*-TOHs, acting as a phosphatase/isomerase, as shown in Scheme 1.5. Here, we report the structure of Rv3378c both in its apo form as well as bound to its TPP substrate, and to a bisphosphonate inhibitor, BPH-629 (Scheme 1.5). We also report the structure of the *M. tuberculosis* *cis*-decaprenyl diphosphate synthase (Rv2361c) with another bisphosphonate inhibitor, BPH-640 (Scheme 1.5) which has a similar structure and inhibitor-binding mode as found with Rv3378c-the phosphatase.

1.6 Drug Discovery for Tuberculosis

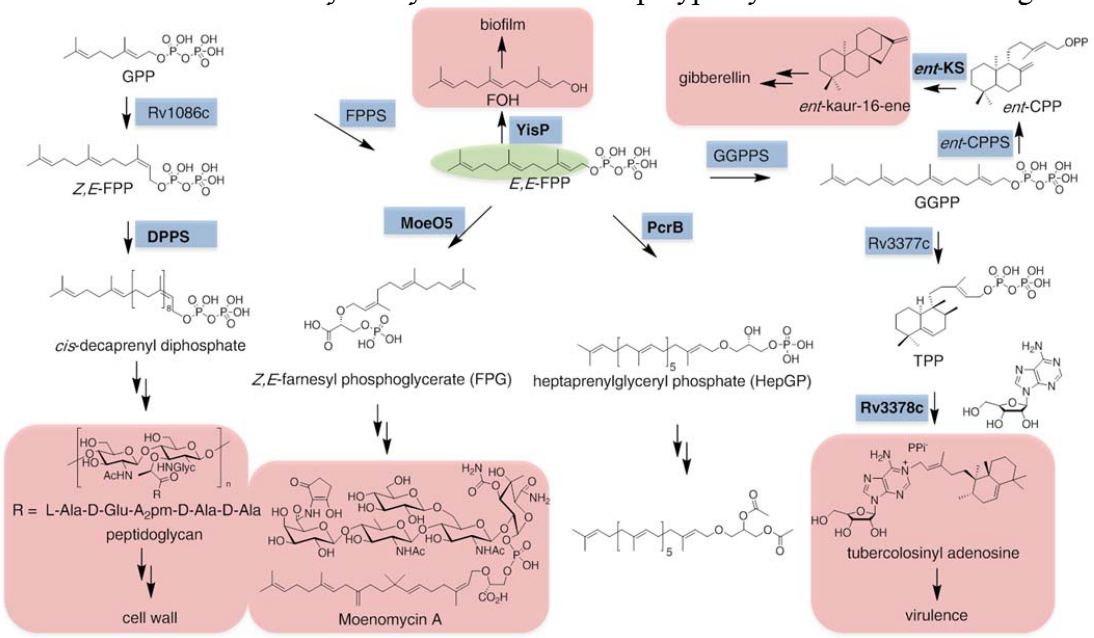
Antibiotic resistance is a public health problem that, arguably, has the potential to destroy the efficacy of all antibiotics in the next 10-20 years[37]. There is, therefore, an urgent need for new drugs-especially ones that might be more "resistance-resistant". In this work, we consider the mechanism(s) of action of the new anti-tuberculosis drug SQ109 (**1**) (Chart 1.1), currently in Phase II clinical trials[43]. This drug candidate appeared of interest since it contains a C_{10} isoprenoid (geranyl) side-chain together with a strongly basic (ethylenediamine) fragment-a likely cationic center-suggesting that it might act as a carbocation isostere for a transition state/reactive intermediate in isoprenoid biosynthesis[44] and, as with other inhibitors of isoprenoid biosynthesis, it might be involved in multi-targeting. **1** was developed in a synthesis/screening program[45] in which ~64,000 ethylenediamine analogs of the anti-tuberculosis drug ethambutol (**2**) were

synthesized. **1** was ~ 4x more active than any of the other leads developed, having a minimum inhibitory concentration (MIC) of ~0.7-1.56 μM against *M. tuberculosis* (H37Rv, Erdman and drug-resistant strains) and insights into its mode of action recently became available when the target of SQ109 was proposed[46] to be MmpL3, a trehalose monomycolate (TMM) transporter, an essential membrane protein that transports TMM into the cell envelope. This conclusion was based on the observation that several *M. tuberculosis* mutants produced via serial passage with several **1**-like inhibitors had mutations in the *mmpL3* gene and cross-resistance to **1**, although these latter effects were rather small[46]. More intriguingly, no spontaneous resistant mutants were obtained when using **1**, suggesting the possibility of multiple targets[46-47]. That idea is supported by the observation that **1** also has activity against other bacteria, e.g. *Helicobacter pylori*[48] as well as against the yeast *Candida albicans*[49], neither of which possess the *mmpL3* gene, suggesting again that other **1** targets are harbored by these organisms, and potentially, by *M. tuberculosis*.

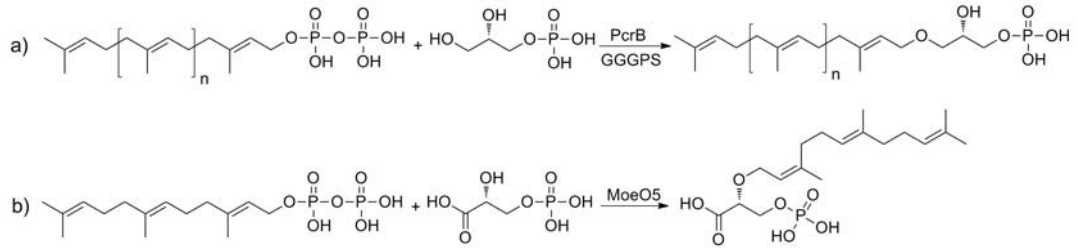
In this work, we synthesized a series of analogs of **1** in which we varied the adamantane head-group and the ethylenediamine linker (varying the possible charge centers), Chart 1.1. All compounds were screened against a panel of bacteria (*M. tuberculosis*, *M. smegmatis*, *Escherichia coli*, *Staphylococcus aureus* and *Bacillus subtilis*), two yeasts (*Saccharomyces cerevisiae* and *Candida albicans*), a malaria parasite (*Plasmodium falciparum*) and a human cell line (MCF-7), to establish anti-bacterial, anti-fungal and anti-malarial structure-activity relationships and to assess mammalian cell toxicity. In addition, we investigated a subset of compounds for activity against a series of putative targets, isoprenoid biosynthesis enzymes, while also investigating the effects of these compounds on respiration, ATP synthesis, and the proton motive force (PMF).

1.7 Schemes, Charts, Tables and Figures

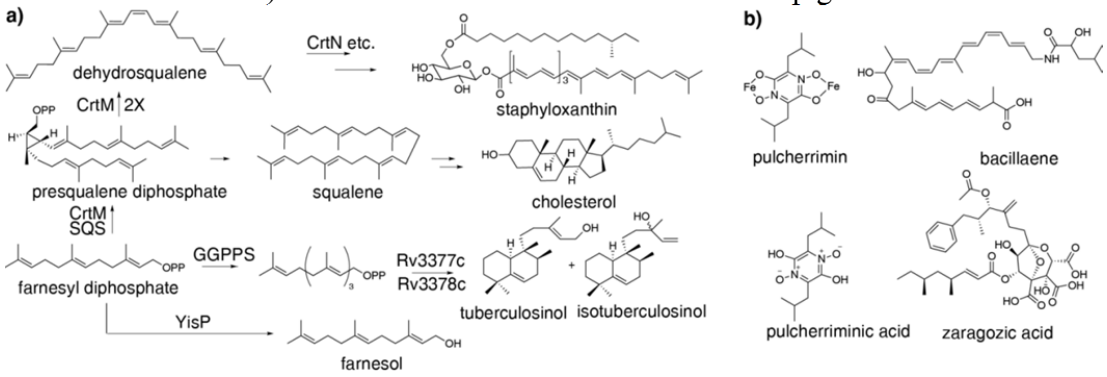
Scheme 1.1 Reactions catalyzed by the six bacterial polyprenyl transferases investigated.



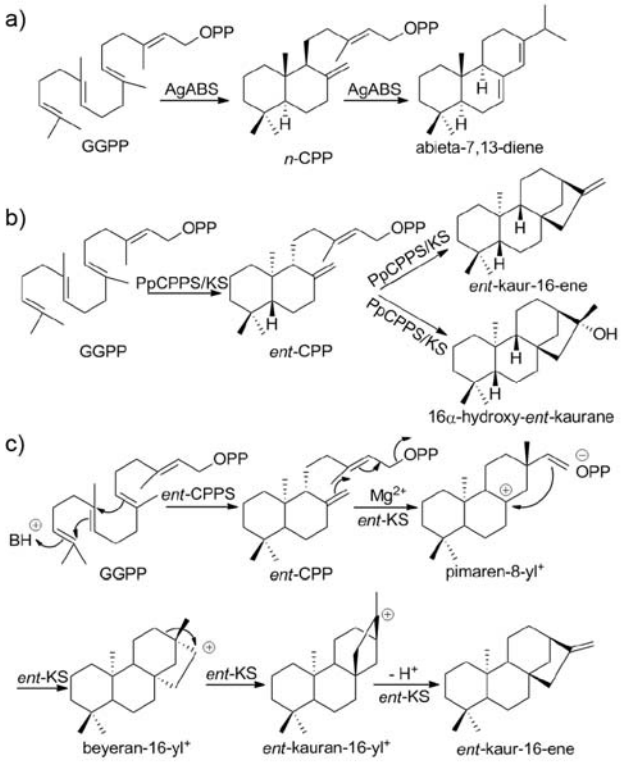
Scheme 1.2 Structures of molecules and reactions of interest. a) Reactions catalyzed by PcrB (n=5) or GGGPS (n=2) b) Reaction catalyzed by MoeO5.



Scheme 1.3 Structures of compounds of interest. a) Biosynthesis of virulence factors, sterols and farnesol. b) A biofilm inhibitor and some *B. subtilis* pigments.



Scheme 1.4 Reactions catalyzed by AgABS, PpCPPS/KS and BjKS, and proposed catalytic mechanism for the BjKS reaction. (a) Abietadiene is formed from GGPP by a bifunctional terpene cyclase abietadiene synthase, in *Abies grandis*. (b) Formation of *ent*-CPP, *ent*-kaurene and kauranol from GGPP catalyzed by *Physcomitrella patens* CPPS/KS. (c) Two proteins are used to produce *ent*-kaurene in the soil bacterium, *Bradyrhizodium japonicum*. Proposed carbocation intermediates are shown.



Scheme 1.5 Biosynthesis of *Mycobacterium tuberculosis* virulence factors (tuberculosinols) and decaprenyl diphosphate, which is essential for cell wall biosynthesis.

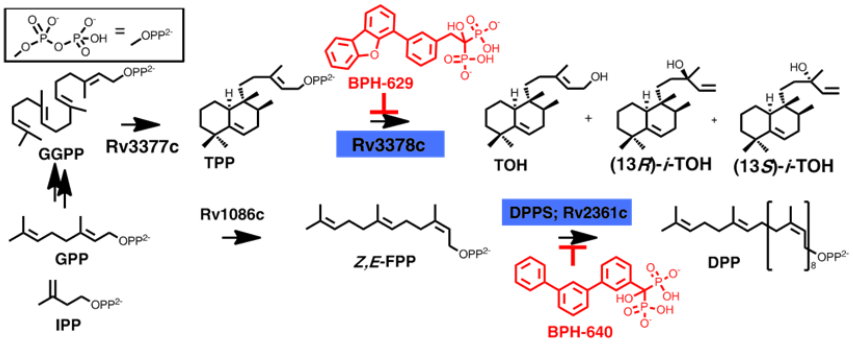


Chart 1.1 Structures of compounds investigated.

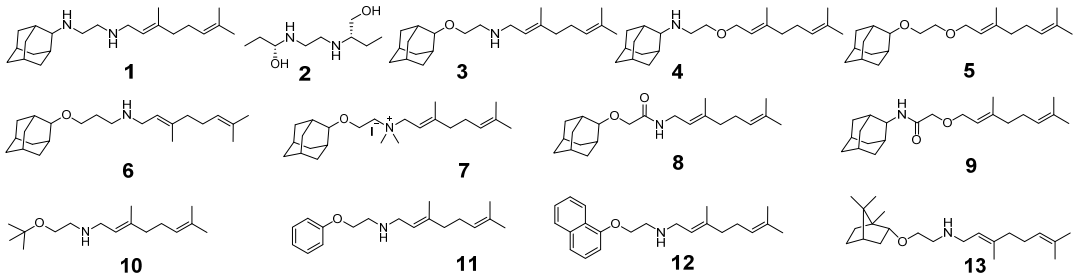


Figure 1.1 Overview of bacterial polyprenyl transferases.

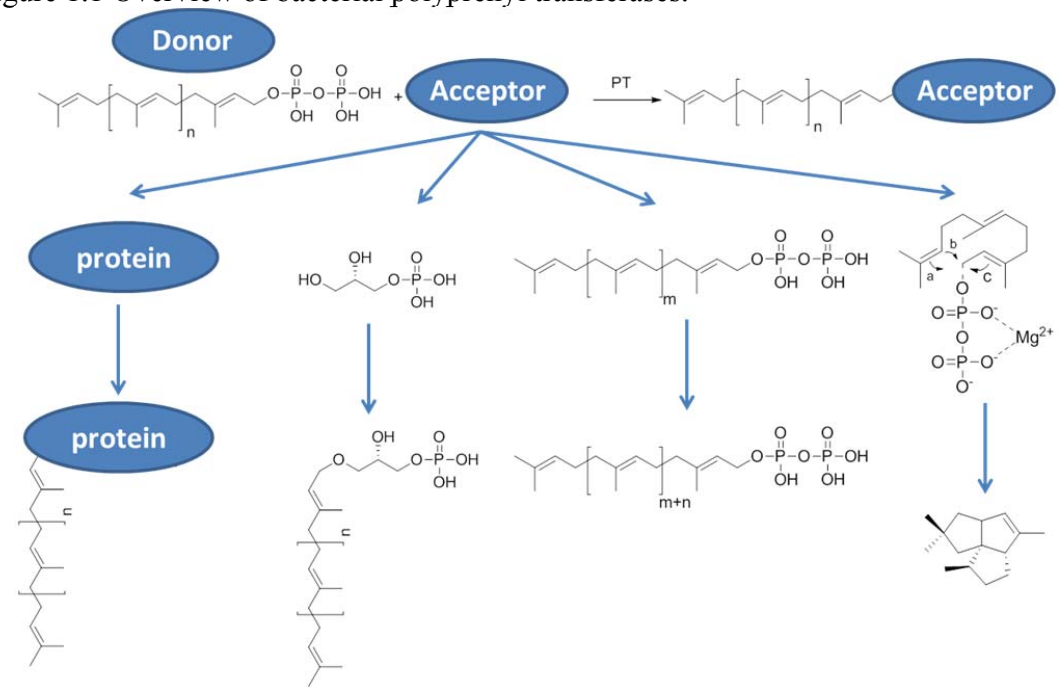


Figure 1.2 Structural phylogenetic tree from bacterial prenyl transferases.

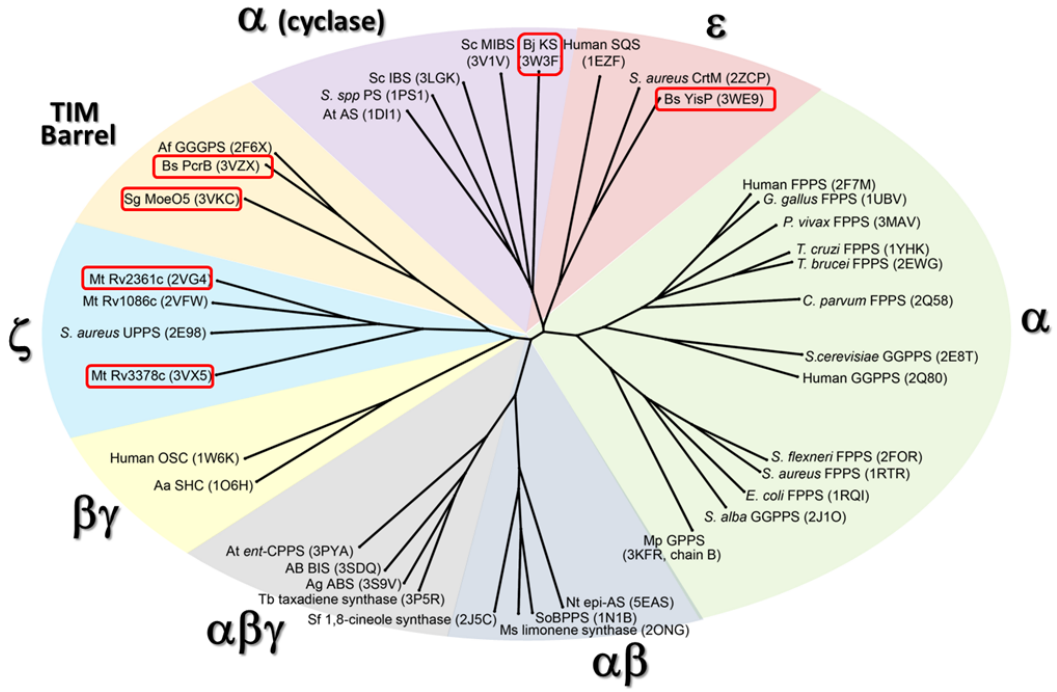
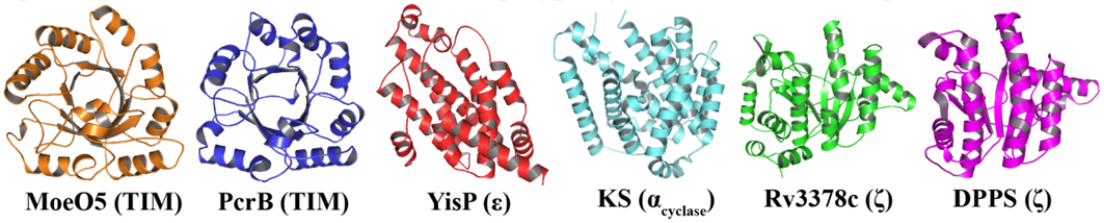


Figure 1.3 Structures of the six bacterial polyprenyl transferases investigated.



1.8 References

1. (a) Casey, P. J.; Seabra, M. C., Protein prenyltransferases. *J. Biol. Chem.* **1996,** *271* (10), 5289-5292; (b) Schafer, W. R.; Rine, J., Protein prenylation - genes, enzymes, targets, and functions. *Annu. Rev. Genet.* **1992,** *26*, 209-237.
2. Payandeh, J.; Fujihashi, M.; Gillon, W.; Pai, E. F., The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J. Biol. Chem.* **2006,** *281* (9), 6070-6078.
3. Matsumi, R.; Atomi, H.; Driessen, A. J. M.; van der Oost, J., Isoprenoid biosynthesis in Archaea - Biochemical and evolutionary implications. *Res. Microbiol.* **2011,** *162* (1), 39-52.
4. Ostash, B.; Walker, S., Moenomycin family antibiotics: chemical synthesis, biosynthesis, and biological activity. *Nat. Prod.Rep.* **2010,** *27* (11), 1594-1617.
5. Buckingham, J., Dictionary of Natural Products on DVD. *CRC, Boca Raton, FL* **2007**.
6. Oldfield, E.; Lin, F. Y., Terpene biosynthesis: modularity rules. *Angew. Chem.* **2012,** *51* (5), 1124-37.
7. (a) Oldfield, E., Targeting isoprenoid biosynthesis for drug discovery: bench to bedside. *Acc. Chem. Res.* **2010,** *43* (9), 1216-26; (b) Cao, R.; Zhang, Y.; Mann, F. M.; Huang, C.; Mukkamala, D.; Hudock, M. P.; Mead, M. E.; Prisic, S.; Wang, K.; Lin, F.-Y.; Chang, T.-K.; Peters, R. J.; Odfield, E., Diterpene cyclases and the nature of the isoprene fold. *Proteins* **2010,** *78* (11), 2417-2432.
8. Ren, F.; Ko, T.-P.; Feng, X.; Huang, C.-H.; Chan, H.-C.; Hu, Y.; Wang, K.; Ma, Y.; Liang, P.-H.; Wang, A. H. J.; Oldfield, E.; Guo, R.-T., insights into the mechanism of the antibiotic-synthesizing enzyme MoeO5 from crystal structures of different complexes. *Angew. Chem.* **2012,** *51* (17), 4157-4160.
9. Ren, F.; Feng, X.; Ko, T.-P.; Huang, C.-H.; Hu, Y.; Chan, H.-C.; Liu, Y.-L.; Wang, K.; Chen, C.-C.; Pang, X.; He, M.; Li, Y.; Oldfield, E.; Guo, R.-T., Insights into TIM-barrel prenyl transferase mechanisms: crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. *Chembiochem* **2013,** *14* (2), 195-199.
10. Xinxin Feng, Y. H., Yingying Zheng, Wei Zhu, Kai Li, Chun-Hsiang Huang, Tzu-Ping Ko, Feifei Ren, Hsiu-Chien Chan, Mulugeta Nega, Shannon Bogue, Daniel López, Roberto Kolter, Friedrich Götz, Rey-Ting Guo, Eric Oldfield, Structure and function of *Bacillus subtilis* YisP: biofilm production by a YisP product. Sumitted to *J. Am. Chem. Soc.* **2014**.
11. Wenting Liu, X. F., Yingying Zheng, Chun-Hsiang Huang, Chiaki Nakano, Tsutomu Hoshino, Shannon Bogue, Tzu-Ping Ko, Chun-Chi Chen, Yunfeng Cui, Jian Li, Eric Oldfield, and Rey-Ting Guo, Structure, function and inhibition of *ent*-kaurene synthase from *Bradyrhizobium japonicum*. Sumitted to *Nature Communication* **2014**.
12. Chan, H.-C.; Feng, X.; Ko, T.-P.; Huang, C.-H.; Hu, Y.; Zheng, Y.; Bogue, S.; Nakano, C.; Hoshino, T.; Zhang, L.; Lv, P.; Liu, W.; Crick, D. C.; Liang, P.-H.; Wang, A. H. J.; Oldfield, E.; Guo, R.-T., Structure and inhibition of tuberculosinol synthase and decaprenyl diphosphate synthase from *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **2014,** *136* (7), 2892-2896.
13. (a) Badger, J.; Sauder, J. M.; Adams, J. M.; Antonysamy, S.; Bain, K.; Bergseid, M. G.; Buchanan, S. G.; Buchanan, M. D.; Batiyenko, Y.; Christopher, J. A.; Emtage, S.; Eroshkina, A.; Feil, I.; Furlong, E. B.; Gajiwala, K. S.; Gao, X.; He, D.; Hendle, J.; Huber,

A.; Hoda, K.; Kearins, P.; Kissinger, C.; Laubert, B.; Lewis, H. A.; Lin, J.; Loomis, K.; Lorimer, D.; Louie, G.; Maletic, M.; Marsh, C. D.; Miller, I.; Molinari, J.; Muller-Dieckmann, H. J.; Newman, J. M.; Noland, B. W.; Pagarigan, B.; Park, F.; Peat, T. S.; Post, K. W.; Radojicic, S.; Ramos, A.; Romero, R.; Rutter, M. E.; Sanderson, W. E.; Schwinn, K. D.; Tresser, J.; Winhoven, J.; Wright, T. A.; Wu, L.; Xu, J.; Harris, T. J. R., Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* **2005,** *60* (4), 787-796; (b) Doud, E. H.; Perlstein, D. L.; Wolpert, M.; Cane, D. E.; Walker, S., Two distinct mechanisms for TIM barrel prenyltransferases in bacteria. *J. Am. Chem. Soc.* **2011,** *133* (5), 1270-1273; (c) Ostash, B.; Doud, E. H.; Lin, C.; Ostash, I.; Perlstein, D. L.; Fuse, S.; Wolpert, M.; Kahne, D.; Walker, S., Complete characterization of the seventeen step Moenomycin biosynthetic pathway. *Biochemistry* **2009,** *48* (37), 8830-8841.

14. Guldan, H.; Matysik, F.-M.; Bocola, M.; Sterner, R.; Babinger, P., Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in bacteria. *Angew. Chem.* **2011,** *50* (35), 8188-8191.

15. Petit, M. A.; Dervyn, E.; Rose, M.; Entian, K. D.; McGovern, S.; Ehrlich, S. D.; Bruand, C., PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol. Microbiol.* **1998,** *29* (1), 261-273.

16. Kourounakis, A. P.; Katselou, M. G.; Matralis, A. N.; Ladopoulou, E. M.; Bavavea, E., Squalene synthase inhibitors: an update on the search for new antihyperlipidemic and antiatherosclerotic agents. *Curr. Med. Chem.* **2011,** *18* (29), 4418-4439.

17. Urbina, J. A., Ergosterol biosynthesis and drug development for Chagas disease. *Mem. Inst. Oswaldo Cruz* **2009,** *104*, 311-318.

18. (a) Wieland, B.; Feil, C.; Gloriamaercker, E.; Thumm, G.; Lechner, M.; Bravo, J. M.; Poralla, K.; Gotz, F., Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of *Staphylococcus-Aureus*. *J. Bacteriol.* **1994,** *176* (24), 7719-7726; (b) Pelz, A.; Wieland, K. P.; Putzbach, K.; Hentschel, P.; Albert, K.; Gotz, F., Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. *J. Biol. Chem.* **2005,** *280* (37), 32493-32498.

19. (a) Liu, G. Y.; Essex, A.; Buchanan, J. T.; Datta, V.; Hoffman, H. M.; Bastian, J. F.; Fierer, J.; Nizet, V., *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* **2005,** *202* (2), 209-215; (b) Clauditz, A.; Resch, A.; Wieland, K. P.; Peschel, A.; Gotz, F., Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect. and Immu.* **2006,** *74* (8), 4950-4953.

20. Lopez, D.; Kolter, R., Functional microdomains in bacterial membranes. *Genes Dev.* **2010,** *24* (17), 1893-1902.

21. Liu, C.-I.; Jeng, W.-Y.; Chang, W.-J.; Ko, T.-P.; Wang, A. H. J., binding modes of zaragozic acid A to human squalene synthase and staphylococcal dehydrosqualene synthase. *J. Biol. Chem.* **2012,** *287* (22), 18750-18757.

22. (a) Dairi, T., Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. *J Antibiot. (Tokyo).* **2005,** *58* (4), 227-243; (b) Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J., Gibberellin biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Lett.* **2009,** *583* (2), 475-480; (c) Hamano, Y.; Kuzuyama, T.; Itoh, N.;

Furihata, K.; Seto, H.; Dairi, T., Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin. *J. Biol. Chem.* **2002,** *277* (40), 37098-37104; (d) Smanski, M. J.; Peterson, R. M.; Shen, B., Platensimycin and platencin biosynthesis in *Streptomyces platensis*, showcasing discovery and characterization of novel bacterial diterpene synthases. In *Meth. Enzymol.*, David, A. H., Ed. Academic Press: **2012**; Vol. 515, pp 163-186; (e) Komakr, H.; Nemoto, A.; Tanaka, Y.; Takgi, H.; Yazawa, K.; Mikami, Y.; Shigemori, H.; Kobayashi, J.; Akikazu, A.; Nagata, Y., Brasilicardin A, a new terpenoid antibiotic from pathogenic *Nocardia brasiliensis*: fermentation, isolation and biological activity. *J Antibiot. (Tokyo).* **1999,** *52* (1), 13-19; (f) Tully, R. E.; van Berkum, P.; Lovins, K. W.; Keister, D. L., Identification and sequencing of a cytochrome P450 gene cluster from *Bradyrhizobium japonicum*. *BBA-Gene. Struct. Expr.* **1998,** *1398* (3), 243-255; (g) Smanski, M. J.; Peterson, R. M.; Huang, S.-X.; Shen, B., Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Curr. Opin. Chem. Biol.* **2012,** *16* (1-2), 132-141; (h) Mann, F. M.; Prisic, S.; Hu, H.; Xu, M.; Coates, R. M.; Peters, R. J., Characterization and inhibition of a class II diterpene cyclase from *Mycobacterium tuberculosis*: implications for tuberculosis. *J. Biol. Chem.* **2009,** *284* (35), 23574-23579; (i) Prach, L.; Kirby, J.; Keasling, J. D.; Alber, T., Diterpene production in *Mycobacterium tuberculosis*. *FEBS J.* **2010,** *277* (17), 3588-3595.

23. (a) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R., Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem.* **2000,** *39* (16), 2812-2833; (b) Wendt, K. U.; Schulz, G. E., Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **1998,** *6* (2), 127-33; (c) Koksal, M.; Hu, H.; Coates, R. M.; Peters, R. J.; Christianson, D. W., Structure and mechanism of the diterpene cyclase *ent*-copalyl diphosphate synthase. *Nat. chem. biol.* **2011,** *7* (7), 431-3.

24. Kawaide, H.; Hayashi, K.-i.; Kawanabe, R.; Sakigi, Y.; Matsuo, A.; Natsume, M.; Nozaki, H., Identification of the single amino acid involved in quenching the *ent*-kauranyl cation by a water molecule in *ent*-kaurene synthase of *Physcomitrella patens*. *FEBS J.***2011,** *278* (1), 123-133.

25. von Schwartzenberg, K.; Schultze, W.; Kassner, H., The moss *Physcomitrella patens* releases a tetracyclic diterpene. *Plant Cell Rep.* **2004,** *22* (10), 780-786.

26. MacMillan, J.; Beale, M. H., Diterpene Biosynthesis. Pergamon Press: 2000.

27. Zhou, K.; Gao, Y.; Hoy, J. A.; Mann, F. M.; Honzatko, R. B.; Peters, R. J., Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from *Abies grandis*. *J. Biol. Chem.* **2012,** *287* (9), 6840-6850.

28. (a) Ernst, J. D., The immunological life cycle of tuberculosis. *Nat. Rev. Immunol.* **2012,** *12* (8), 581-591; (b) Lienhardt, C.; Glaziou, P.; Uplekar, M.; Lonnroth, K.; Getahun, H.; Raviglione, M., Global tuberculosis control: lessons learnt and future prospects. *Nat. Rev. Microbiol.* **2012,** *10* (6), 407-16; (c) Zumla, A.; Hafner, R.; Lienhardt, C.; Hoelscher, M.; Nunn, A., Advancing the development of tuberculosis therapy. *Nat. Rev. Drug. Discov.* **2012,** *11* (3), 171-2.

29. Koul, A.; Arnoult, E.; Lounis, N.; Guillemont, J.; Andries, K., The challenge of new drug discovery for tuberculosis. *Nature* **2011,** *469* (7331), 483-90.

30. (a) Liu, G. Y.; Essex, A.; Buchanan, J. T.; Datta, V.; Hoffman, H. M.; Bastian, J. F.; Fierer, J.; Nizet, V., *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* **2005,** *202* (2), 209-

15; (b) Liu, C. I.; Liu, G. Y.; Song, Y.; Yin, F.; Hensler, M. E.; Jeng, W. Y.; Nizet, V.; Wang, A. H.; Oldfield, E., A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* **2008,** *319* (5868), 1391-4.
31. (a) Nakano, C.; Hoshino, T., Characterization of the Rv3377c gene product, a type-B diterpene cyclase, from the *Mycobacterium tuberculosis* H37 genome. *Chembiochem* **2009,** *10* (12), 2060-71; (b) Hoshino, T.; Nakano, C.; Ootsuka, T.; Shinohara, Y.; Hara, T., Substrate specificity of Rv3378c, an enzyme from *Mycobacterium tuberculosis*, and the inhibitory activity of the bicyclic diterpenoids against macrophage phagocytosis. *Org. Biomol. Chem.* **2011,** *9* (7), 2156-65; (c) Nakano, C.; Ootsuka, T.; Takayama, K.; Mitsui, T.; Sato, T.; Hoshino, T., Characterization of the Rv3378c gene product, a new diterpene synthase for producing tuberculosinol and (13R, S)-isotuberculosinol (nosyberkol), from the *Mycobacterium tuberculosis* H37Rv genome. *Biosci. Biotechnol. Biochem.* **2011,** *75* (1), 75-81.
32. Rudi, A.; Aknin, M.; Gaydou, E.; Kashman, Y., Asmarines I, J, and K and nosyberkol: four new compounds from the marine sponge *Raspailia sp*. *J. Nat. Prod.* **2004,** *67* (11), 1932-5.
33. (a) Maugel, N.; Mann, F. M.; Hillwig, M. L.; Peters, R. J.; Snider, B. B., Synthesis of (+/-)-nosyberkol (isotuberculosinol, revised structure of edaxadiene) and (+/-)-tuberculosinol. *Org. Lett.* **2010,** *12* (11), 2626-9; (b) Spangler, J. E.; Carson, C. A.; Sorensen, E. J., Synthesis enables a structural revision of the *Mycobacterium tuberculosis*-produced diterpene, edaxadiene. *Chem. Sci.* **2010,** *1* (2), 202-205.
34. Pethe, K.; Swenson, D. L.; Alonso, S.; Anderson, J.; Wang, C.; Russell, D. G., Isolation of *Mycobacterium tuberculosis* mutants defective in the arrest of phagosome maturation. *Proc. Natl. Acad. Sci. U S A* **2004,** *101* (37), 13642-7.
35. Mann, F. M.; Xu, M.; Chen, X.; Fulton, D. B.; Russell, D. G.; Peters, R. J., Edaxadiene: a new bioactive diterpene from *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **2009,** *131* (48), 17526-7.
36. Mann, F. M.; Prisic, S.; Hu, H.; Xu, M.; Coates, R. M.; Peters, R. J., Characterization and inhibition of a class II diterpene cyclase from *Mycobacterium tuberculosis*: implications for tuberculosis. *J. Biol. Chem.* **2009,** *284* (35), 23574-9.
37. (a) US Centers for Disease Control (2013) Antibiotic Resistance Threats in the United States, 2013. http://www.cdc.gov/features/AntibioticResistanceThreats/index.htlm; (b) World Health Organization. http://www.who.int/tb/publications/global_report/en/index.html.
38. Report of Two Workshops on Novel Antimirobial Therapeutics Council, N. R., Ed. The National Academies Press: Washintgon DC, 2006.
39. Gray, K. C.; Palacios, D. S.; Dailey, I.; Endo, M. M.; Uno, B. E.; Wilcock, B. C.; Burke, M. D., Amphotericin primarily kills yeast by simply binding ergosterol. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (7), 2234-9.
40. Fischbach, M. A., Combination therapies for combating antimicrobial resistance. *Curr. Opin. Microbiol.* **2011,** *14* (5), 519-23.
41. Silver, L. L., Multi-targeting by monotherapeutic antibacterials. *Nat. Rev. Drug Discov.* **2007,** *6* (2), 126-126.
42. Morphy, J. R., The Challenges of Multi-Target Lead Optimization. In *Designing Multi-Target Drugs*, 2012; pp 141-154.

43. Sacksteder, K. A.; Protopopova, M.; Barry, C. E.; Andries, K.; Nacy, C. A., Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action. *Future Microbiol.* **2012,** *7* (7), 823-837.
44. Martin, M. B.; Arnold, W.; Heath, H. T.; Urbina, J. A.; Oldfield, E., Nitrogen-containing bisphosphonates as carbocation transition state analogs for isoprenoid biosynthesis. *Biochem. Bioph. Res. Co.* **1999,** *263* (3), 754-758.
45. Protopopova, M.; Hanrahan, C.; Nikonenko, B.; Samala, R.; Chen, P.; Gearhart, J.; Einck, L.; Nacy, C. A., Identification of a new antitubercular drug candidate, SQ109, from a combinatorial library of 1,2-ethylenediamines. *J. Antimicrob. Chemother.* **2005,** *56* (5), 968-974.
46. Tahlan, K.; Wilson, R.; Kastrinsky, D. B.; Arora, K.; Nair, V.; Fischer, E.; Barnes, S. W.; Walker, J. R.; Alland, D.; Barry, C. E.; Boshoff, H. I., SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. *Antimicrob. Agents. Ch.* **2012,** *56* (4), 1797-1809.
47. La Rosa, V.; Poce, G.; Canseco, J. O.; Buroni, S.; Pasca, M. R.; Biava, M.; Raju, R. M.; Porretta, G. C.; Alfonso, S.; Battilocchio, C.; Javid, B.; Sorrentino, F.; Ioerger, T. R.; Sacchettini, J. C.; Manetti, F.; Botta, M.; De Logu, A.; Rubin, E. J.; De Rossi, E., MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. *Antimicrob. Agents Chemother.* **2012,** *56* (1), 324-331.
48. Makobongo, M. O.; Einck, L.; Peek, R. M.; Merrell, D. S., In vitro characterization of the anti-bacterial activity of SQ109 against *Helicobacter pylori*. *Plos One* **2013,** *8* (7).
49. Barbosa, e. a., In vitro antifungal susceptibility testing of drug candidate SQ109 *Candida albicans*. In *Intersicence Conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, San Francisco, CA, 2006.

Chapter 2: Insights into the Mechanism of the Antibiotic-Synthesizing Enzyme MoeO5 from Crystal Structures of Different Complexes

2.1 Notes and Acknowledgements

C.-H. Huang, H. -C. Chan and Y. Hu cloned, expressed and purified MoeO5 proteins. F. Ren, T.-P. Ko and R.-T. Guo carried out crystallization, data collection, structure determination and refinement. X. Feng obtained *cis*-farneysl pyrophosphate via biosynthesis and studied kinetics of MoeO5. K. Wang synthesized *trans*-farnesyl pyrophosphate and farnesyl thiopyrophosphate. Y. Ma, P.-H. Liang, A. H.-J. Wang, E. Oldfield and R.-T. Guo analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by National Basic Research Program of China (grant 2011CB710800 to RTG), Tianjin Municipal Science and Technology Commission (10ZCKFSY06000 to RTG), and the National Institutes of Health (AI074233 to EO). We thank the National Synchrotron Radiation Research Center of Taiwan for beam-time allocation and data-collection assistance.

This chapter was reproduced in part with permission from F. Ren, T.-P. Ko, X. Feng, C.-H.Huang, H.-C. Chan, Y. Hu, K. Wang, Y. Ma, P.-H. Liang, A. H.-J. Wang, E. Oldfield and R.-T. Guo. *Angew. Chem. Int. Ed .Engl*. Volume 51, Issue 17, page 4240, April 23, 2012.

2.2 Introduction

The phosphoglycolipid antibiotic moenomycin directly blocks bacterial cell wall biosynthesis by inhibiting peptidoglycan glycosytransferases.[1] The enzyme MoeO5, encoded by the *moe* gene cluster 1 in *Streptomyces ghanaensis*, catalyzes the initial step of moenomycin production in which the C_{15}-hydrocarbon moiety of farnesyl pyrophosphate (FPP) is transferred to the 2-hydroxyl group of 3-phosphoglycerate (3PG), forming an ether bond (Figure 2.1).[2] The reaction is similar to that of geranylgeranylglyceryl phosphate synthase (GGGPS) for synthesizing Archaea-type phospholipids.[3] However, unlike GGGPS whose product retains the all-*trans*

configuration of the isoprenyl chain, the farnesyl group transferred by MoeO5 undergoes a *trans*-to-*cis* isomerization at the C2-C3 double bond.[4] The crystal structures of GGGPS and the bacterial homologue PcrB reveal a triose phosphate isomerase (TIM)-barrel fold, which had not been observed previously in prenyltransferases (PTs).[3, 5] The sequences of GGGPS and PcrB share 35% amino acid identity, but MoeO5 shows only 10% identity to both enzymes (Figure 2.2). Here we report the X-ray crystallographic structures of MoeO5 bound to the product 2-*Z*,*E*-farnesyl-3-phosphoglycerate (FPG), a substrate analogue farnesyl thiopyrophosphate (FsPP), magnesium ion (Mg^{2+}), and pyrophosphate ion (PPi) that, together with additional biochemical and bioinformatics results, shed light on the possible mechanisms of action of this unusual enzyme.

2.3 Results and Discussions

Molecular-replacement (MR) approaches to determine the MoeO5 structure by using GGGPS and PcrB as search models were not successful, reflecting perhaps the significant variations in the protein sequences. Because MoeO5 contains no Cys residue, to solve the structure by using multiple isomorphous replacement (MIR), we produced the mutant H97C for efficient preparation of mercury-based MIR derivatives (Table 2.1). The other structures were solved by MR (Figure 2.3 and Table 2.2). See the Supporting Information (SI) for details. MoeO5 crystallizes as a homodimer (Figure 2.4 and Table 2.3). The dimer interface buries 1200 $Å^2$, or more than 10% surface area, on each monomer and mainly involves hydrophobic residues in helices α4 and α5. The *cis*-peptide of Pro141 binds to a Mg^{2+} ion at the molecular dyad (Figure 2.4). These helices also mediate dimerization in GGGPS and PcrB, but here one of the TIM barrels is rotated by 180° (Figure 2.5). More description of the protein structure as well as the bound ligands can be found in the SI. Despite its Cα deviation of about 2.0 Å from the GGGPS and PcrB monomer (Figure 2.6), the similar protein fold with a connecting loop (denoted λ3) between strands β3 and β4 clearly places MoeO5 among this new class of TIM-barrel PTs.

The wild-type MoeO5 co-crystallized with a bound product FPG, of which the C_{15} tail moiety makes a U-turn mainly at the C10-C11 bond (Figure 2.7), rather than extending straight into a nonpolar groove along helix α4 as supposed in GGGPS and

PcrB.[3,5] The equivalent groove in MoeO5 is obstructed by the λ3 loop (Figure 2.8), which contains His97 and is four-residues longer than in GGGPS and PcrB (Figure 2.2).[4] This loop may act as a swinging door for binding and enclosure of the prenyl substrate in GGGPS,[3] but the door is virtually locked up in MoeO5 and the resulting small volume seems barely able to accommodate a C_{10} group. However, substitution of a Tyr residue of GGGPS and PcrB by Ala157 in strand β5 makes a "nook" for the bent C_{15} group in MoeO5 (Figure 2.8). Neither the C_{10} GPP nor the C_{20} GGPP is a substrate of MoeO5, and soaking with the thio-analogue GGsPP has no effect on the bound FPG. Clearly the cavity is specific for FPP, and the bent conformation is likely to be important in precisely positioning the C_{15} group for the *trans*-to-*cis* conversion.

On the other hand, the 3PG head group binding site is equivalent to that for glycerol 1-phosphate (G1P) in GGGPS (Figure 2.9).[3] The 3'-phosphate, sandwiched between loop β7-α7 and helix α8', is fastened by three backbone NH groups (Figure 2.10a). The 1'-carboxyl group, together with His97, also binds to a water molecule. When the crystals were soaked with FsPP, partial substitution of FPG occurred in three hours, and it was mostly replaced overnight. The β-phosphate group of FsPP occupies the same position as does the 3'-phosphate of FPG, but the α-phosphate is loosely bound (Figure 2.10b). Each active site also shows a Mg-PPi complex adjacent to the FsPP, which may represent an alternative disposition of the PPi moiety of FsPP (Figure 2.11). Here, the Mg^{2+} ion binds to Asp41, which is highly conserved in MoeO5, GGGPS and PcrB. A similar site containing Mg^{2+} and PO_4^{3-} was found with a 3PG soak, but PPi alone did not bind to this site (detailed in the SI). Like most other PTs, MoeO5 requires Mg^{2+} for activity and the network of interactions with Mg^{2+} is expected to facilitate FPP ionization, the first step in catalysis.

In GGGPS and PcrB the prenyl carbocation is directly attached to the C3'-hydroxyl of G1P, and the *trans*-configuration is conserved (Figure 2.12a). However, in MoeO5 Doud et al. recently proposed catalytic mechanisms in which either nerolidyl pyrophosphate (NPP) was formed after ionization, facilitating bond rotation and the *trans*-to-*cis* conversion (Figure 2.12b), or that *Z*,*E*-FPP was an intermediate.[4] The observation that *trans*-FPP and NPP react at the same rate while *cis*-FPP reacts ~5x faster (Figure 2.13) suggests that all three species may be involved. Another possibility is that

the His97-bound water molecule, which is 3.6 Å from the C3 atom, can be activated and associate with the farnesyl cation, transiently forming nerolidol (NOH), facilitating bond rotation (Figure 2.12c), and we do find that H97C is an inactive mutant, supporting a vital role of His97, although **R**,**S**-*trans*-NOH (+ 3PG + PPi + Mg^{2+}) is not a substrate.

2.4 Conclusions

In summary, MoeO5 forms a TIM barrel structure and binds FPG in a curved pocket, mainly as a result of its long λ3 loop. An FPP ionization site containing Asp41 and Mg^{2+} is located nearby, and the results obtained here are consistent with formation of a *Z*,*E*-FPP intermediate. We also find that His97 in the λ3 loop is essential for activity. Except for using Asp41 to bind Mg-PPi, the *cis*-bond formation catalyzed by MoeO5 is distinct from those of other *cis*-PTs.[6] To further study this novel mechanism by mutagenesis, kinetic analysis and other physicochemical measurements, the various ligand-complex structures presented here provide a good starting point.

2.5 Materials and Methods

Expression and purification of MoeO5

The gene encoding MoeO5 from *Streptomyces ghanaensis* was synthesized chemically and amplified by polymerase chain reaction (PCR) with forward primer 5'-GGTATTGAGGGTCGCGCTGGTGCTGGTGCTATGAACGCCAGCCCTCAACTGG ATCAT-3' and reverse primer 5'-AGAGGAGAGTTAGAGCCATTAGCGACCGCTACCTGGACGAGCTGG-3', and then cloned into the pET32 Xa/LIC vector. Because MoeO5 does not contain any Cys residue, for efficient preparation of mercury-containing derivatives for multiple isomorphous replacement (MIR) phasing, we made the H97C mutant near the C-terminus of strand β3 (Figure 2.2). Based on the sequence alignment, this residue should be accessible to the solvent but not too flexible for a heavy atom site. The H97C mutant was prepared with MoeO5–pET32a Xa/LIC vector as the template and a forward primer of 5'-CCGCTGCCTGTCGTACTGTGCTTCCCGCCGCGCCCTGGT-3' was used. The recombinant plasmids were transformed to *E. coli* BL21 (DE3) and the protein was induced with 0.4 mM isopropyl-thiogalactopyranoside (IPTG) at 20 °C for 24 hours.

Cell paste was harvested by centrifugation at 7,000 x *g* and resuspended in a lysis buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM NaCl and 20 mM imidazole. Cell lysate was prepared with a French Press Instrument, and then centrifuged at 17,000 x *g* to remove cell debris. The proteins were purified by FPLC using a Ni-NTA column. The buffer and gradient for the Ni-NTA column were 25 mM Tris, pH 7.5, 150 mM NaCl, and 20–250 mM imidazole. His-tagged MoeO5 was eluted at about 110 mM imidazole. The protein solution was dialyzed against a buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM NaCl and then subjected to Factor Xa digestion to remove the His tag. The mixture was then passed through Ni-NTA column again and the untagged MoeO5 was eluted with 20 mM imidazole-containing buffer. The protein was then purified by FPLC using DEAE column, the buffer and gradient for the DEAE column were 25 mM Tris, pH 7.5, and 0–500 mM NaCl. The protein was eluted at about 225 mM NaCl. The purified protein was finally concentrated to 5 mg mL^{-1} in a 25 mM Tris, 150 mM NaCl, pH 7.5 buffer. The preparation procedure of H97C mutant was almost identical to that of the wild-type protein.

Crystallization, data collection, structure determination and refinement

The wild-type MoeO5 protein was first crystallized by using the PEG/Ion 2 screen kit (Hampton Research) and sitting-drop vapor diffusion method. The reservoir solution (No. 6) contained 0.2 M sodium malonate, pH 6.0, and 20% w/v PEG3350. Better crystals were obtained by optimizing the reservoir composition, which contained 0.2 M sodium malonate, pH 6.0, and 22% PEG3350. The reservoir for crystallizing the H97C mutant was slightly different; it contained 0.2 M sodium malonate, pH 6.0 and 26% PEG3350. All crystals were prepared at room temperatures. They reached suitable sizes for X-ray diffraction in 2 days. Various ligands were used in soaking the wild-type crystals before data collection. The "native" crystal had actually been soaked in 5 mM GGsPP, 10 mM $MgCl_2$, 23% PEG 3350, 0.2 M sodium malonate, pH 6.0, for 3 hours. The first FsPP-soaked (for 3 hours) crystal was prepared using 20 mM FsPP, 10 mM $MgCl_2$, 23% PEG 3350, 0.2 M sodium malonate, pH 6.0. The second FsPP-soaked (overnight) crystal used 100 mM instead of 20 mM FsPP in the solution, whereas the PPi-soaked and 3PG-soaked (both for 3 hours) crystals used 100 mM sodium pyrophosphate

and 20 mM 3PG, respectively. For heavy-atom derivatives, the Hg-containing reagents of Heavy Atom Screen Hg (Hampton Research) were used. Cryoprotectant solutions (0.2 M sodium malonate, pH 6.0, 28% w/v PEG3350 and 10% glycerol) containing 2 mM Hg derivatives were used in soaking the H97C mutant crystals for 3 to 5 hours.

The X-ray diffraction data sets were collected at beam lines BL13B1 and BL13C1 of the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan) and processed by using the program of HKL2000.[7] Prior to structural refinement, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor. The MIR datasets of mercury-containing derivatives were collected at a wavelength of 0.9763 Å (BL13C1). Using SOLVE and RESOLVE,[8] combinations of datasets from different Hg-derivative crystals with the "native" dataset from the wild-type crystal improved the figure of merit (FOM) values from 0.46 to 0.51, the Z-scores from 16.4 to 40.1, and the number of auto-built amino acid residues up to 384. The best results were obtained using the derivatives of ethylmercurithiosalicylic acid, mercury potassium iodide, ethylmercury chloride, and tetrakis(acetoxymercuri)methane. Statistics of the four Hg-derivative data sets and the MIR phasing are summarized in Table 2.1. Because there are two MoeO5 molecules in an asymmetric unit of the monoclinic crystal, a more continuous electron density map was constructed by superimposing all of the initial crude models, and a more complete model was manually reconstructed. The model and map were further improved by computational refinement using Refmac5.[9] Finally a complete model (still missing the N-terminal residues) with most side chains was built by ARP/wARP.[10] Structures of the triclinic crystals were solved by molecular replacement (MR) using a monomeric model from the monoclinic crystal and the program CNS.[11] It turned out that each crystal contains the same dimeric arrangement of two MoeO5 molecules. All of the following structural refinements were carried out using the programs CNS and Coot.[12] The electron densities for the active site bound ligands are shown in Figure 2.3. Data collection and refinement statistics of these crystals are summarized in Table 2.2. All diagrams of the protein structures were prepared by using the software PyMOL (http://www.pymol.org/).

Description of the overall protein structure

The monomer of MoeO5 folds into a TIM barrel, but the helix α3 connecting the two adjacent strands β3 and β4 in most other TIM-barrel structures is replaced by a loop that lacks significant secondary structure formation. This loop is denoted λ3 as shown in the sequence alignment diagram of Figure 2.2 here and the molecular diagram of Figure 2.7. As observed in GGGPS, PcrB, and many other TIM-barrel proteins, the N-terminal helix α0 located on the N-terminal side of the parallel β strands seals the bottom of the barrel. before the three helices α4, α5 and α8, three associated helices denoted α4', α5' and α8' are found on the C-terminal side, or top of the barrel. In all crystal forms of MoeO5, an asymmetric unit contains a homodimer (Figure 2.4). The dimer interface of MoeO5 is comparable to those of GGGPS and PcrB, buries an average of 1200 $Å^2$ surface area, and accounts for more than 10% of the solvent-accessible surface on each monomer. It involves a number of hydrophobic interactions contributed by Leu121, Tyr127, Val129, Trp130, Phe133, Leu134, Leu137, Ala138, Phe140, Val189, Ala192, Phe193 and Phe195 from helices α4 and α5, as well as Leu154 from strand β5 and the N-terminal Trp20. Two salt bridges are formed by the side chains of Arg21 and Glu135*. (Residues from the other monomer are denoted by asterisks.) The side chain of Lys131 is hydrogen bonded to the carbonyl groups of Phe193* and Gly194*, in addition to the salt bridge to Asp126. The flat guanidinium group of Arg185 stacks with the imidazole group of His188*. Moreover, near the C-terminus of helix α4, the carbonyl groups of Leu137, Ala138 and Phe140, together with their equivalents in the dyad-related monomer, form a six-coordinated complex with a Mg^{2+} ion at the dimer interface. This Mg^{2+} is not found in GGGPS or PcrB. Interestingly, as shown in Figure 2.5, the two monomers in a MoeO5 homodimer are arranged differently than those in the GGGPS and PcrB dimers. Although the same helices α4 and α5 form the dimer interface, the dyad axis relating the two MoeO5 monomers is roughly perpendicular to that relating the GGGPS or PcrB monomers. Consequently, MoeO5 has its two TIM barrels facing the opposite sides, whereas GGGPS and PcrB have theirs facing the same side. The monomer of MoeO5 also deviates more from those of GGGPS and PcrB than between the latter two (Figure 2.6).

Consistent with the canonical rule for a TIM-barrel protein, the active site of MoeO5 is located on the C-terminal end of the parallel β-strands that form a funnel-

shaped structure, or the top of the barrel. As shown in Figure 2.8, the product FPG and the Mg-PPi complex are both bound to this site, which contains a deep pocket. A citrate and a glycerol 1-phosphate are also located in the equivalent binding site of GGGPS (Figure 2.9). The shape and size of the cavities in MoeO5 and GGGPS are different. In GGGPS the tunnel-like cavity, or the nonpolar slide, is long and presumably can accommodate a C_{20}-geranylgeranyl chain in an extended conformation. By contrast, the equivalent cavity in MoeO5 is blocked half-way mainly because of the longer λ3 loop. On the other hand, the substitution of Tyr124 in GGGPS by Ala157 in MoeO5 created a "nook" for the C_{15}-farnesyl group, which is bound in a curved manner. Despite the different crystal forms and the presence of different bound ligands, the protein conformation of MoeO5 remains unchanged. The dimer models of MoeO5 superimpose well on one another, either in a straight forward way or with the two monomers swapped (Table 2.3). A unique *cis*-peptide of Phe140-Pro141, consistently observed in all crystals, binds to a Mg^{2+} ion at the dimer interface and this should be important in maintaining the protein structure. In the refined model, the peptide dihedral angles in a Ramachandran plot showed a few outliers. These include Leu19, Trp130 and Glu205. Electron density of the N-terminal Leu19 is very clear. It is in the second position of a type II' turn,[13] where the carbonyl C=O group of Pro18 is hydrogen bonded to the amide NH of Arg21. The conformation of Trp130 is also well defined. It is located in a theoretically allowed region to the far left side of the α-helical region. This Trp130 is at the junction of helices α4' and α4 and heavily involved in the dimer interface. Glu205 is located on the protein surface, but its NH and C=O groups are bound to the side chains of Ser202 and Asn230, which in turn are bound to the C=O of Phe160 and Arg203, respectively.

Description of ligand structures in the native MoeO5 and after FsPP, PPi and 3PG soaking

In the MoeO5 crystal a bound product FPG was observed in each active site, probably because both substrates should be present in significant levels inside the *E. coli* cells when the protein was expressed. 3PG is a central metabolite in glycolysis, and FPP is also widely utilized in the various cellular activities. The product does not dissociate easily, and presumably would require the presence of other downstream enzymes for

direct transfer, because FPG is a secondary metabolite for antibiotics production. As mentioned above, soaking with GGsPP had little effect on the bound FPG molecule. The larger C_{20} geranylgeranyl chain is unable to fit into the cavity that is specific for binding to the C_{15} farnesyl chain. However, soaking with FsPP turned out to be effective in replacing the bound product, although it occurred gradually. The PPi moiety of FsPP is bound at the same site as is the 3PG moiety of FPG. The presence of Mg^{2+} in the FsPP soaking condition also allowed the binding mode of the Mg-PPi complex to be observed. The bound FsPP to the 3PG site, however, does not have a Mg^{2+} ion associated with the PPi, and consequently the 3PG site is not specific for the PPi-moiety of FPP. In fact, the FsPP overnight-soak structure can be alternatively refined with the PPi moiety of FsPP swapped with that of the Mg-PPi complex (Figure 2.11a). This model showed acceptable geometry and yielded R and R_{free} values of 0.162 and 0.191 using the same data set. However, judging by the electron density, a majority of PPi in the FsPP is still bound to the 3PG site.

The enzyme-ligand interactions are shown in Figure 2.10 of the main text. In the Mg-PPi binding site, the six-coordinated Mg^{2+} ion is chelated by a water molecule and the side chains of Asp41, Ser69 and Asp71 in addition to two oxygen atoms of the PPi. The side chains of Lys44, Thr70, Arg165 and Arg203, together with the backbone NH group of Thr70, make several hydrogen bonds to the PPi. Although not charged, the shorter and more rigid side chain of Thr70 can be important in properly orienting the substrate molecule for the catalyzed reaction. Extensive soaking experiments with PPi showed that PPi alone did not bind to this site, but to a more remote site between two protein molecules at lattice contact. The level of PPi in the solution was 10-fold higher than the level of Mg^{2+}, and probably Mg^{2+} was mostly chelated by two or three molecules of PPi and consequently was depleted. (We also found salt crystals in the same drop.) Soaking with 3PG also had little effect on the tightly bound FPG, but showed a bound Mg-phosphate in the Mg-PPi binding site. Interestingly, in the heavy-atom free H97C mutant crystals, the active site contained an unknown ligand with a longer hydrocarbon moiety that is also bent. Substitution of the imidazole group by a sulfhydryl enlarged the cavity. A phosphate group was seen in the equivalent position as was that in 3PG. However, it seemed to be attached to a carbonyl group, which made a hydrogen bond to

the side chain of Ser228. Although the ligand size appeared to resemble the Archaea-type lipid GGGP, the electron density did not suggest such an identity. It was modeled as a 1-phospho-palmitic acid, which contains a C_{15} tail in addition to the head group, and the structure refined to R and R_{free} values of 0.169 and 0.211 at 1.82 Å. The disposition of the hydrocarbon tail is quite different from that of FPG (Figure 2.11b). Whatever the ligand is, these observations suggest that the mutant H97C is no longer capable of synthesizing FPG, and the *trans*-to-*cis* conversion machinery for the farnesyl group is not functional.

Preparation of *Z*,*E*-FPP with Rv1086 and MoeO5 activity assay

Rv1086 was expressed and purified as previously described.[14] To make *Z*,*E*-FPP, 1 μM Rv1086 was added into 10 ml 50 mM Tris-HCl buffer (pH 7.2) containing 300 μM GPP, 1 mM IPP, 1 mM $MgCl_2$ and 300 mM NaCl. The mixture was incubated at 37 °C for overnight. After the reaction, Rv1086 and precipitated *Z*,*E*-FPP was removed by passing the reaction mixture through an Amicon centrifuge tube (10 kDa). *Z*,*E*-FPP concentration was determined by LC/MS with a standard curve generated with *E*,*E*-FPP. LC/MS analyses were carried out using an Agilent LC/MSD Trap XCT Plus instrument. Compounds were separated on a 5 μm (4.6 x 150 mm) Eclipse XDB-C8 column (Agilent) using a 0 – 100% acetonitrile (in 25 mM NH_4HCO_3 buffer with 0.1% triethylamine) gradient, and monitored by using negative-ion mode ESI.

Activity of MoeO5 was measured as previously described.[15] In brief, the assays were performed in 50 mM Tris-HCl at pH 7.2, 300 mM NaCl, 1mM $MgCl_2$ with 1 mM 3-**D**-phosphoglyceric acid, and 0.5 μM MoeO5. *E*,*E*-FPP or *Z*,*E*-FPP was added at a saturated concentration (200 μM for *E*,*E*-FPP and 91 μM for *Z*,*E*-FPP). The reactions were incubated at 37°C, quenched with an equal volume of methanol and then analyzed by LC/MS. The results are shown in Figure 2.13. Other substrates NPP, NOH, GPP and GGPP were also tested for MoeO5 activity under similar conditions.

Analysis of the TIM-barrel PT sequences

In order to assess the likely importance of different residues in catalysis, we used the Scorecons program to rank order residues in terms of their essentiality.[16] We used as input a JPRED3 alignment based on the *S. ghanaenesis* MoeO5 sequence,[17] then

ordered the Scorecons results based on the scores obtained. The most essential residue in MoeO5 is Asp41 with a score of 0.904.

A possible catalytic mechanism that involves NOH

The bound ligands to the wild-type MoeO5 in the refined crystal structures are superimposed in Figure 2.14. Based on the ligand arrangements and the measured distances, the Mg-PPi is 3.0 Å and 3.6 Å from the C1 atom of FPG and FsPP, but at least 5 Å from C3. Thus it is unlikely for the PPi to move to the C3 atom, forming a NPP, but the PPi tends to associate with C1 of the intermediate to form *Z,E*-FPP under proper conditions. A possible catalytic mechanism is thus proposed. The reaction starts by sequential binding to 3PG and FPP, the latter substrate FPP should be in the form of Mg^{2+} complex in order to interact properly with Asp41, Ser69, Thr70 and Asp71, as well as the positively charged Lys44, Arg165 and Arg203. Asp41, important in Mg-PPi binding, is highly conserved in the TIM barrel PTs. The imidazole side chain of His97 as a general acid-base then activates the water molecule by subtracting a proton from it. The water then attacks at the C3 atom of FPP, forming NOH. The double-bonded group of C1=C2 then rotates about the single bond C2-C3 and converts the conformation of the intermediate from a *trans*-like conformation to a *cis*-like conformation. The rotation is a small change in term of the molecular structure, and can be assisted by the nearby Mg-PPi complex, which is a good leaving group but still likely to re-associate with the intermediate to form *Z,E*-FPP. Meanwhile, the 3-OH group of NOH obtains a proton from His97, and the farnesyl group re-develops a positively charged nature among the C1-C2-C3 bonds. Glu167 in GGGPS has been suspected to act as a general base to polarize and subtract a proton from the 3'-OH group of G1P,[3] but the equivalent Tyr201 in MoeO5 is not positioned in a proper orientation for the same role. Instead, the adjacent Mg-PPi can serve to accept a proton from the 2'-OH group of 3PG, which in turn forms a new bond with the *Z,E*-farnesyl group. The resulting FPG remains in the active-site pocket of MoeO5, whereas the Mg-PPi seems not so tightly bound. Other possibilities exist, including the possibility that in cells, *Z,E*-FPP is used directly, but then the observation that FPP, NPP and *Z,E*-FPP all react at not dissimilar rates would have to be coincidental, rather than indicative of an ionization/isomerization mechanism.

2.6 Schemes, Charts, Tables and Figures

Table 2.1 Summary of Hg-derived protein data collection and phasing statistics.

Name	H97C-native	H97C-$C_9H_9HgNaO_2S$	H97C-K_2HgI_4	H97C-C_2H_5HgCl	H97C-$C(HgOOCCH_3)_4$
Data collection					
Wavelength (Å)	0.9763	0.9763	0.9763	0.9763	0.9763
Resolution (Å)	25-1.82 (1.89-1.82)	25-1.49 (1.54-1.49)	25-1.83 (1.90-1.83)	25-1.75 (1.81-1.75)	25-1.71 (1.77-1.71)
Space group			$P2_1$		
Unit-cell					
a (Å)	58.80	58.43	58.39	58.12	58.27
b (Å)	84.57	84.62	84.48	84.04	84.31
c (Å)	59.48	59.52	59.46	59.41	59.35
β (°)	112.01	111.75	112.03	111.72	111.68
No. of measured reflections	153907 (15347)	496607 (49447)	201166 (19089)	170748 (16972)	175642 (17270)
No. of unique reflections	48096 (4796)	85622 (8381)	46783 (4656)	53359 (5304)	56659 (5571)
Completeness (%)	99.8 (100.0)	97.6 (95.9)	99.6 (99.9)	99.7 (99.6)	98.5 (97.6)
R_{merge} (%)[a]	7.8 (50.0)	6.0 (39.6)	6.0 (22.2)	6.6 (48.2)	5.0 (14.3)
Mean *I*/ (*I*)	18.9 (3.4)	29.1 (6.0)	21.5 (8.4)	18.4 (3.5)	21.5 (11.5)
Multiplicity	3.2 (3.2)	5.8 (5.9)	4.3 (4.1)	3.2 (3.2)	3.1 (3.1)
Phasing					
No. of sites			4		
Z-score			40.06		
Figure of merit			0.51		

Values in parentheses are for the highest resolution shell.

[a] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl)-\langle I(hkl)\rangle| \sum_{hkl}\sum_i I_i(hkl)$.

Table 2.2 Data collection and refinement statistics for the MoeO5 crystals. All positive reflections were used in the refinement. Values in parentheses are for the outermost resolution shells.

	Native	FsPP (3 hours)	FsPP (overnight)	PPi	3PG
Data collection					
Space group	$P2_1$	$P1$	$P1$	$P1$	$P1$
Unit-cell					
***a* (Å)**	59.0	46.6	46.7	46.7	46.8
***b* (Å)**	84.5	58.9	58.6	59.7	58.7
***c* (Å)**	59.6	58.8	58.9	59.1	59.2
α (°)	90.0	97.6	97.7	66.7	97.8
β (°)	112.3	108.4	112.2	71.1	112.6
γ (°)	90.0	112.5	108.6	66.0	108.2
Resolution (Å)	25 – 1.39 (1.44 – 1.39)	25 – 1.57 (1.63 – 1.57)	25 – 1.8 (1.86 – 1.80)	25 – 1.66 (1.72 – 1.66)	25 – 1.66 (1.72 – 1.66)
Unique reflections	105467 (10291)	70651 (6964)	46862 (4623)	59540 (5880)	59984 (5912)
Redundancy	5.1 (5.0)	4.0 (4.0)	4.0 (3.9)	4.0 (4.0)	4.0 (4.0)
Completeness (%)	97.5 (95.5)	96.8 (95.1)	96.6 (95.2)	96.1 (94.8)	96.2 (95.0)
Average I/σ(I)	24.8 (3.4)	42.4 (6.4)	37.3 (9.0)	37.1 (9.4)	39.9 (11.4)
R_{merge} (%)	6.9 (55.6)	3.3 (26.6)	3.2 (9.9)	3.1 (9.8)	3.5 (8.6)
Refinement					
No. of reflections	99738 (8573)	69049 (6216)	46545 (4495)	59066 (5627)	59609 (5666)
R_{work} (95% of data)	0.168 (0.222)	0.163 (0.203)	0.161 (0.179)	0.161 (0.184)	0.158 (0.179)
R_{free} (5% of data)	0.189 (0.250)	0.191 (0.243)	0.192 (0.209)	0.186 (0.227)	0.187 (0.220)
R.m.s.d. bonds (Å)	0.020	0.019	0.020	0.020	0.020
R.m.s.d. angles (°)	1.9	1.9	2.0	1.9	2.0
Dihedral angles					
Most favored (%)	96.2	95.8	95.8	96.2	96.0
Allowed (%)	2.6	3.0	3.2	2.8	2.8
Disallowed (%)	1.2	1.2	1.0	1.0	1.2
No. of non-H atoms					
Protein	3927	3927	3943	3932	3927
Water	590	467	405	432	470
Ligand	53	53	72	62	59
Average B (Å^2)					
Protein	13.1	18.6	18.0	13.9	14.1
Water	26.9	31.6	28.6	26.1	26.4
Ligand	12.6	21.2	25.3	17.6	15.8
PDB ID code	***3VK5***	***3VKA***	***3VKB***	***3VKC***	***3VKD***

Table 2.3 Deviations between the MoeO5 models.
The root-mean-square deviations were calculated by superimposing 2000 backbone atoms in a dimer and expressed in Å. Values along the diagonal are for the self comparison by superimposing chain A on chain B and vice versa. Values in the upper triangle are calculated by direct superposition of the different models, and those in the lower triangle are for cross comparison by superimposing chain A on chain B and vice versa.

	Native	**FsPP (3h)**	**FsPP (ON)**	**PPi**	**3PG**
Native	**0.237**	0.162	0.276	0.159	0.131
FsPP (3h)	0.229	**0.179**	0.184	0.138	0.107
FsPP (ON)	0.287	0.224	**0.169**	0.242	0.219
PPi	0.224	0.216	0.267	**0.197**	0.154
3PG	0.215	0.198	0.254	0.218	**0.178**

Figure 2.1 The first step in moenomycin biosynthesis. The reaction is catalyzed by MoeO5, which transfers the C_{15} farnesyl group from FPP to 3PG. The carbon atoms are numbered 1 – 15 in FPP and 1' – 3' in 3PG. Importantly, the original all-*trans* configuration is converted to 2-*cis*,6-*trans* (*Z*,*E*) upon the farnesyl transfer, forming FPG and PPi.

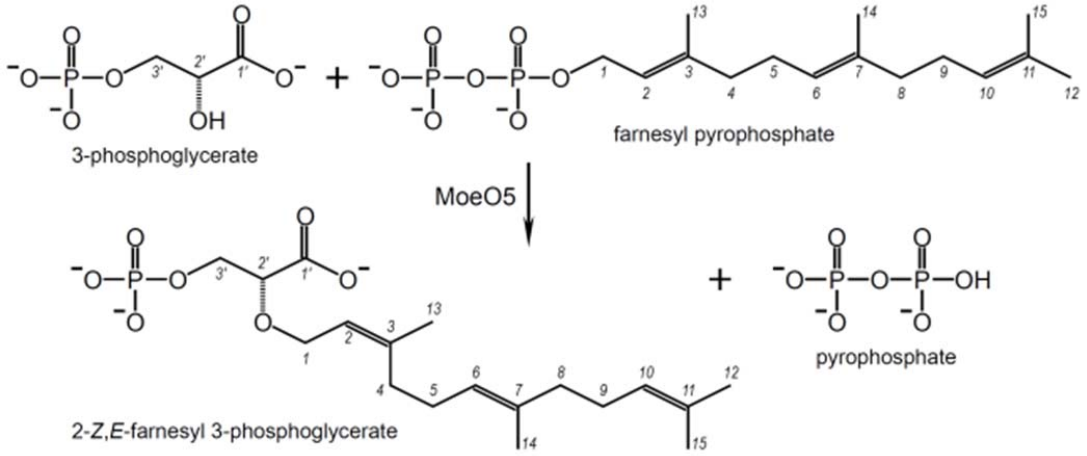


Figure 2.2 Sequences of the TIM-barrel PTs. The amino-acid sequence of MoeO5 is aligned with those of GGGPS and PcrB based on their three-dimensional structures. The cyan bars and green arrows at the bottom denote the locations of α-helices and β-strands, respectively, in the MoeO5 structure. The λ3 loop which replaces the helix α3 commonly observed in other TIM-barrel proteins is represented as a pink bar. Identical amino acids in any two proteins are highlighted in yellow, but marked orange when they occur in all three proteins. The numbers on the top are for MoeO5.

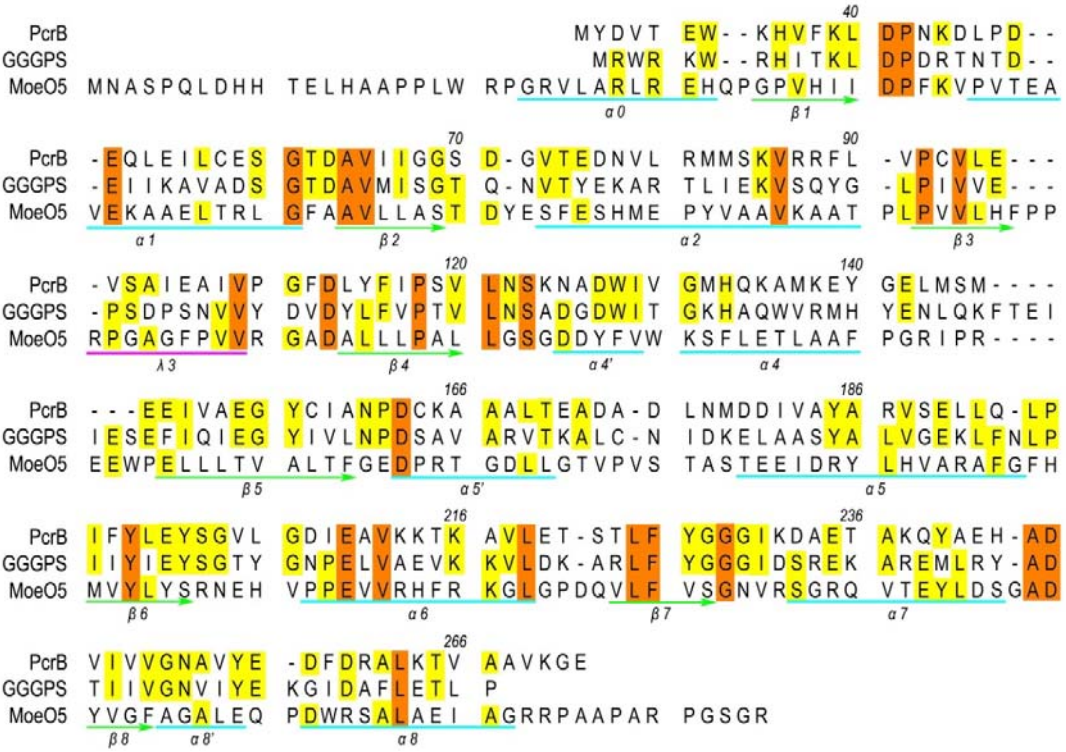
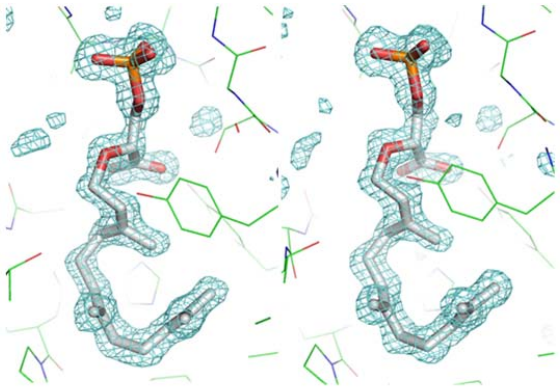
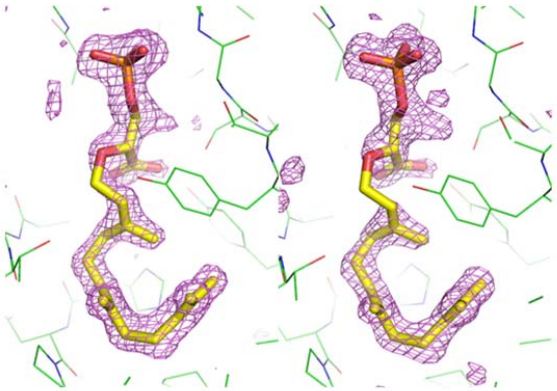


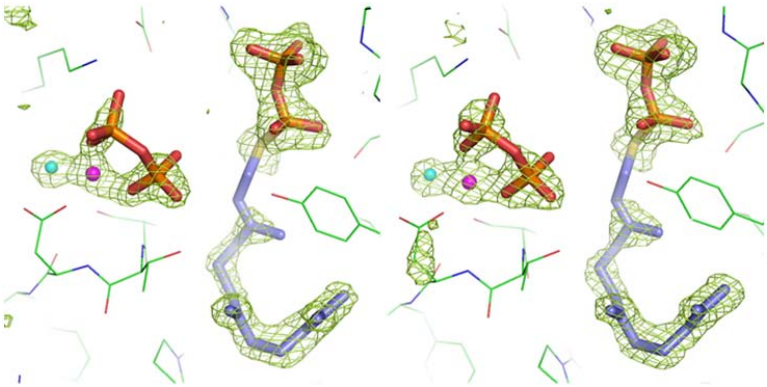
Figure 2.3 Electron densities for the bound ligands. The refined models are superimposed on the composite omit maps calculated using CNS. The maps are contured at 3-σ level and shown as mesh representations. The last two pictures are for the mutant H97C, of which the structure was refined to R_{work} / R_{free} = 0.169 / 0.211 at 1.82-Å resolution. Each of the difference Fourier maps with scaled Fo-Fc coefficients were calculated by combining 20 "omit models" each had the target atoms removed and was then subjected to simulated-annealing at 2500 K in order to remove model bias. The omitted models are shown as thick sticks and balls in the figures, while the remaining parts are shown as thin sticks.



Native

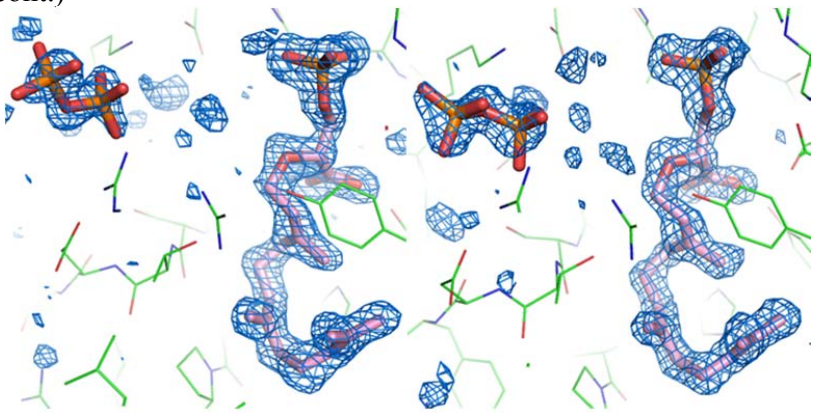


FsPP (3hours)

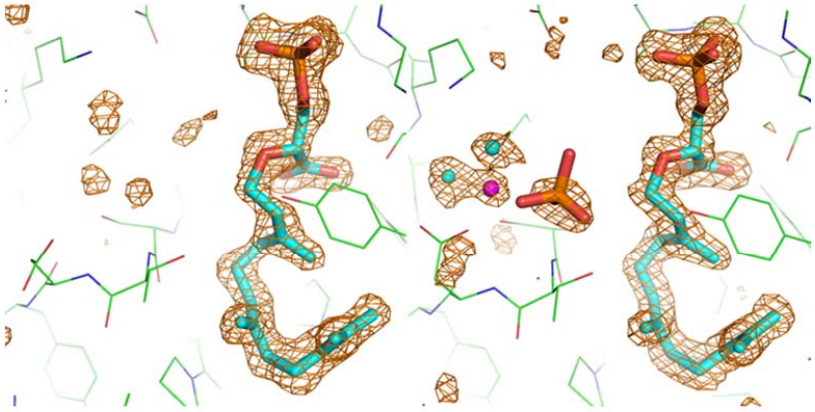


FsPP (overnight)

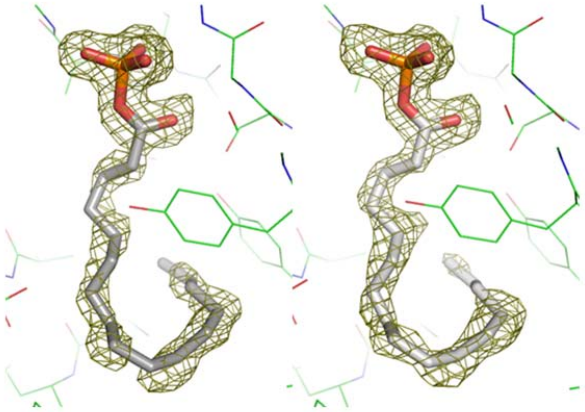
Figure 2.3 (cont.)



PPi



3PG



H97C

Figure 2.4 Structure of the MoeO5 dimer. One monomer is shown as a ribbons diagram and the other monomer as a surface representation. Not found in GGGPS and PcrB, a Mg^{2+} ion is bound at the dimer interface of MoeO5. The coordination involves the *cis*-peptide of Phe140-Pro141, as well as the backbone carbonyl groups of Leu137 and Ala138.

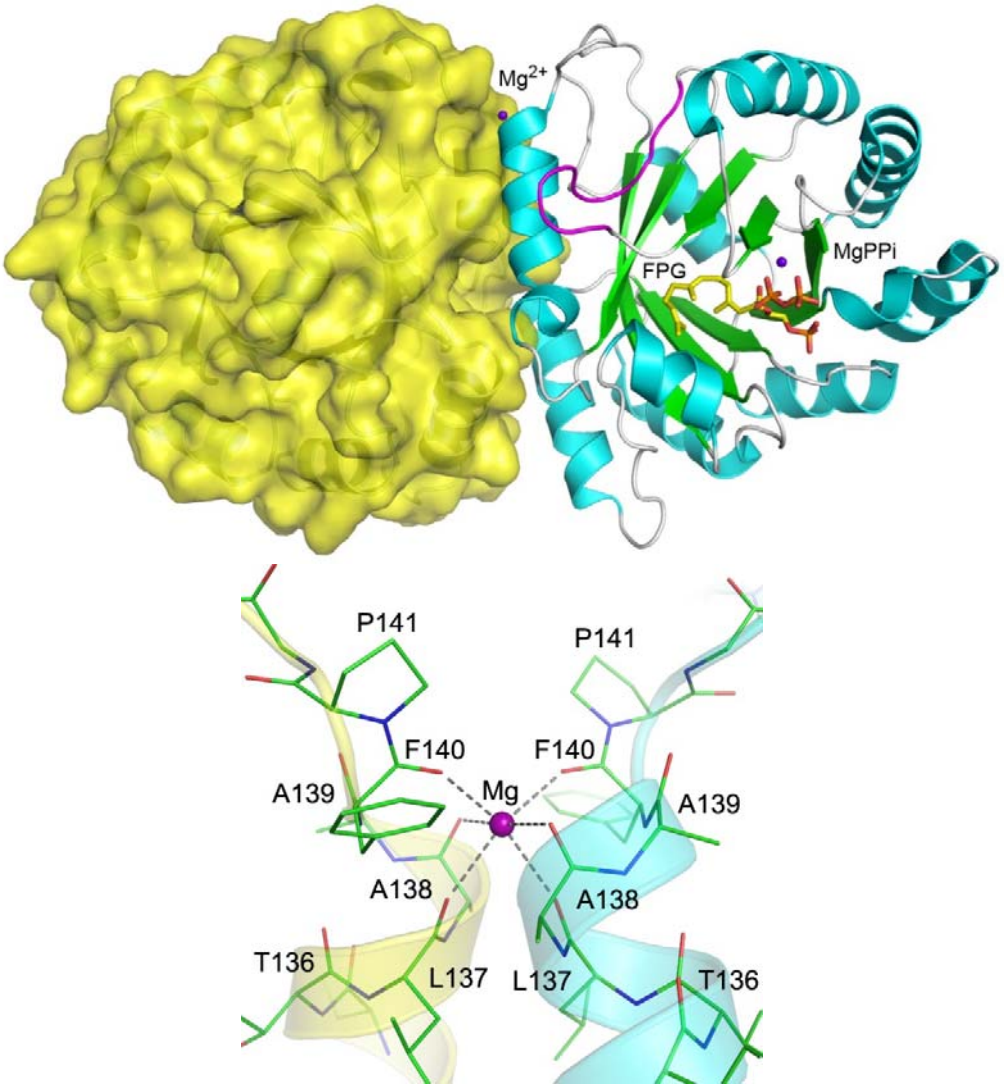


Figure 2.5 Difference of the MoeO5 dimer from two other dimers. The dimers of GGGPS, PcrB and MoeO5 are shown in two orthogonal views. Although the overall shapes of the dimers seem to be similar, the dyad axis which relates the two MoeO5 monomers is disposed differently from those of GGGPS and PcrB. In the dimers shown on the top of the figure, for example, the dyad axes of GGGPS and PcrB are perpendicular to the plane of the page, but that of MoeO5 lies inside the plane. The two different ways of dimer arrangement can be described by two cups. In GGGPS and PcrB, the opening of the TIM barrels are both facing the same side, whereas in MoeO5 they are facing the opposite sides.

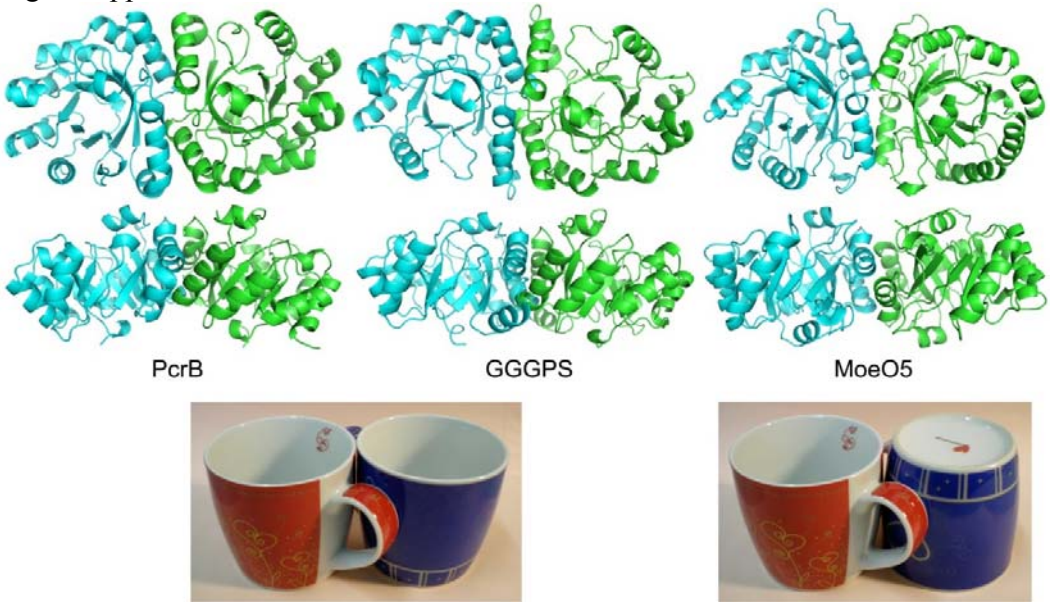


Figure 2.6 Structure comparison of the TIM-barrel PTs. The monomers of MoeO5, GGGPS and PcrB are superimposed and shown as worm diagrams colored in red, green and blue, respectively. The root-mean-square deviations (rmsd) between the matched numbers of Cα atoms are also indicated, as well as the numbers of identical residues. Note that the GGGPS and PcrB structures are much more similar to each other than to MoeO5. The additional diagram at the bottom shows that the dimers of GGGPS and PcrB also superimpose very well. However, the MoeO5 dimer has a different arrangement and cannot be superimposed as a whole on either of the other two dimers.

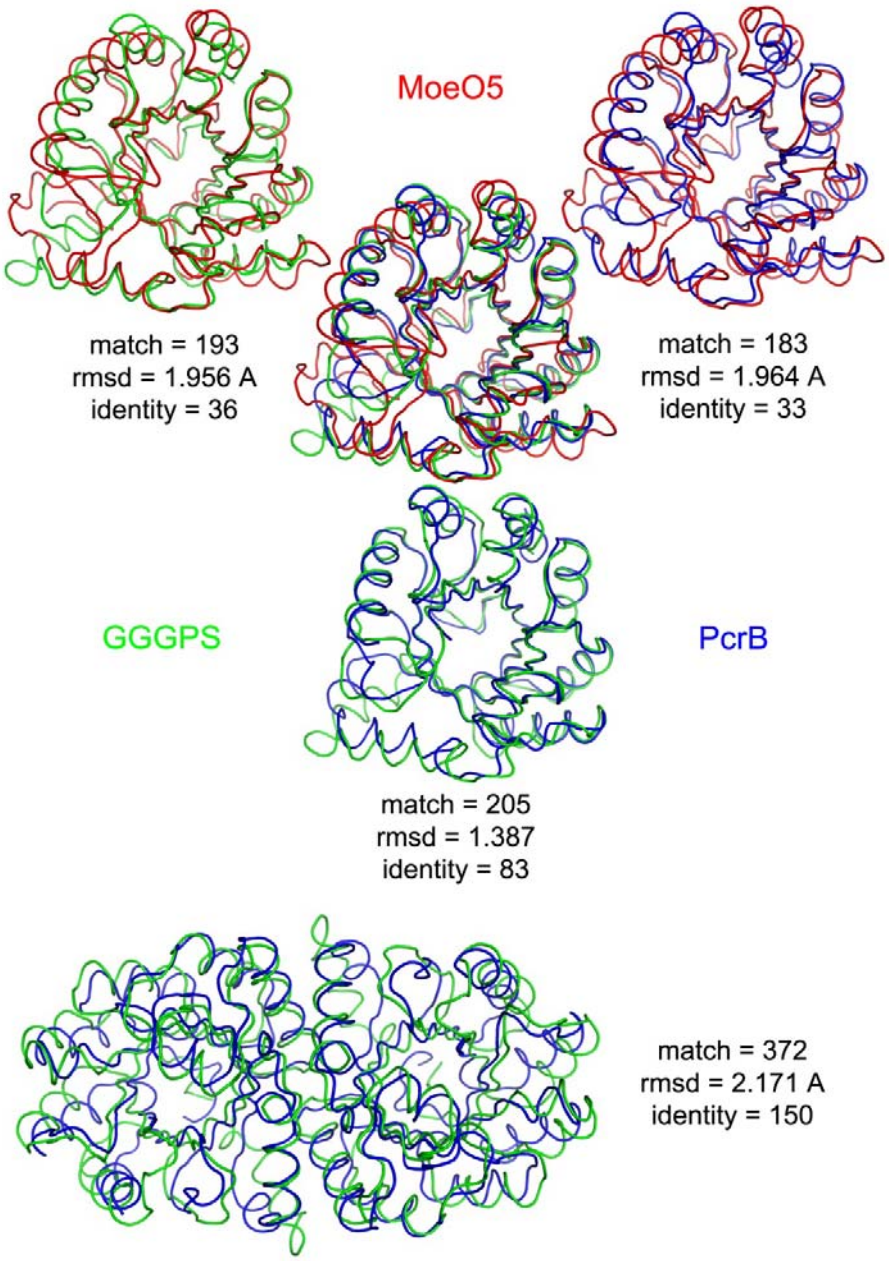


Figure 2.7 Structure of the MoeO5 monomer. The central eight-stranded β-barrel and the surrounding α-helices (numbered 0 – 8) in the ribbon diagram are colored green and cyan. The unusual loop of λ3 is highlighted in magenta. The bound FPG and Mg-PPi complex (from the FsPP-soak crystal) are depicted as ball-and-stick models. A fraction of the other monomer in the MoeO5 dimer is also shown, as a yellow surface presentation.

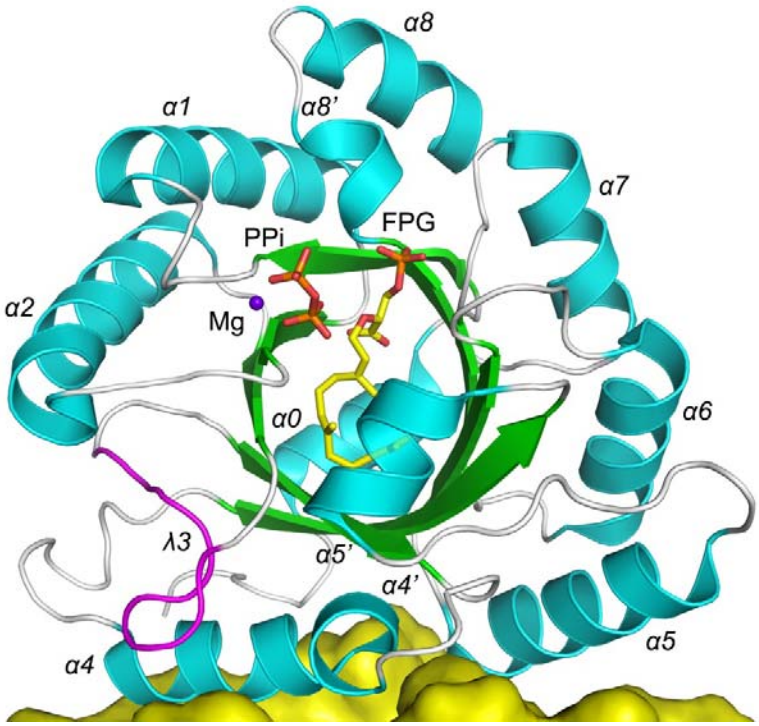


Figure 2.8 Comparison of the prenyl substrate binding pockets. The monomers of MoeO5 (red), GGGPS (green) and PcrB (blue) are superimposed and shown as worm models in the upper panel. The side chains in the λ3 loop and the bound FPG (yellow) are shown as stick models. In the lower panels the surface representations of the enzymes are superimposed on the stick models. The ligand shown here is the product FPG from MoeO5. Here the hydrocarbon-binding tunnels of MoeO5 and GGGPS have different shapes. The Mg-PPi binding site and the unique "nook" in MoeO5 are indicated, as well as the residues Pro99, Pro100, Phe106 and Leu120 that block the "hydrophobic slide", which is supposed to accommodate the longer prenyl chain in GGGPS. Some equivalent residues in GGGPS are also labeled.

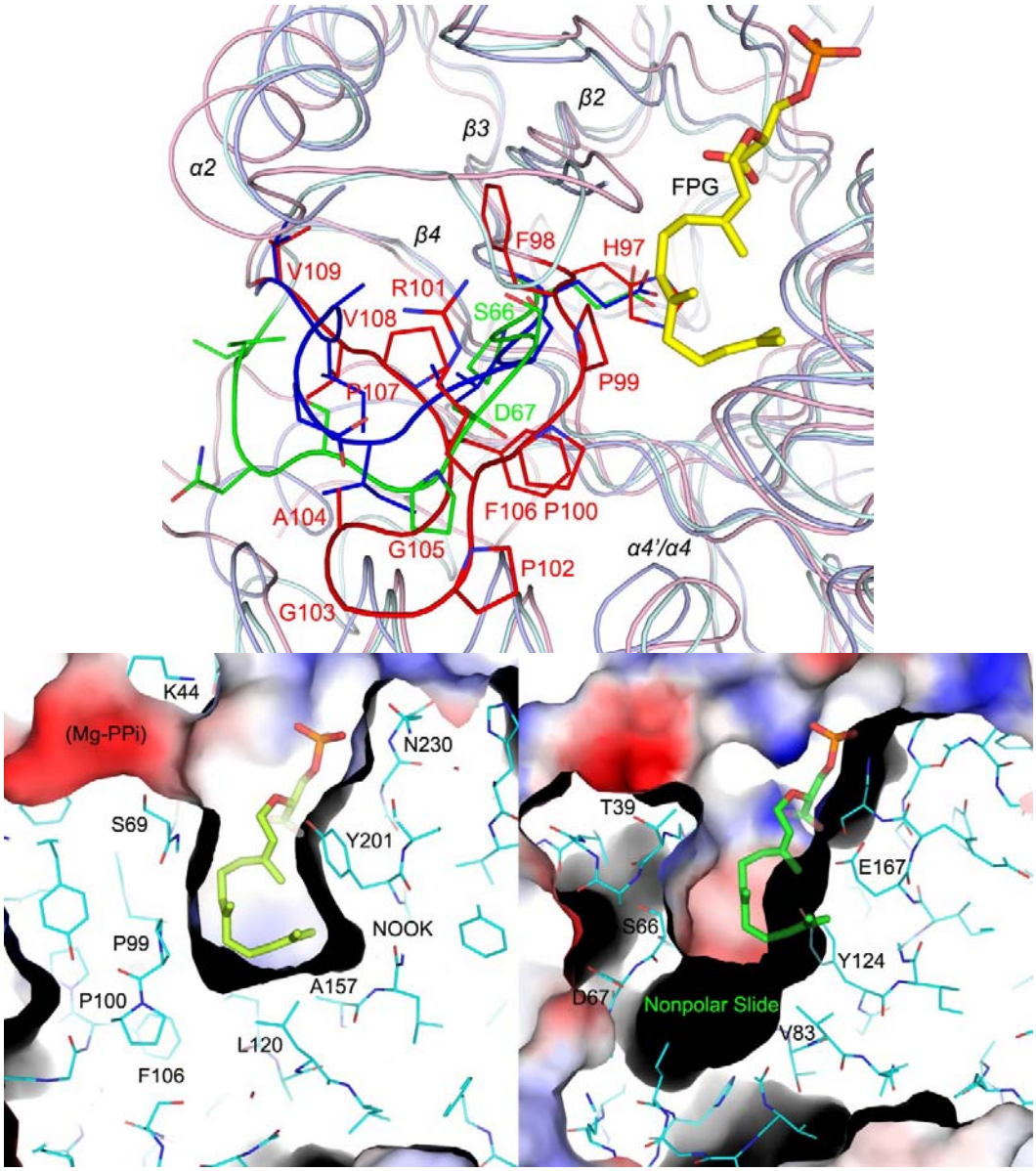


Figure 2.9 Comparison of the bound acceptor substrates. The solvent accessible molecular surfaces of the monomers are viewed from the "top" of the TIM barrel, revealing a deep substrate binding pocket in the active site. The surface is color-coded from red to white and from white to blue according to the electrostatic potentials from -65 k_BT to 0 and from 0 to +65 k_BT. On the left the "native" MoeO5 protein contains a bound product FPG, which is shown as a stick model. The Mg-PPi complex from the FsPP-complex crystal is also shown. In the active site of GGGPS, a bound citrate molecule is shown (PDB 2F6U). The pocket of GGGPS appears to be deeper and more open than that of MoeO5. In the lower panel the FPG from MoeO5 (yellow carbons) and the G1P from GGGPS (PDB 2F6X; cyan carbons), after protein superposition focusing on the two phosphate-binding loops, show an giving an rmsd of 0.42 Å for eight matched atoms.

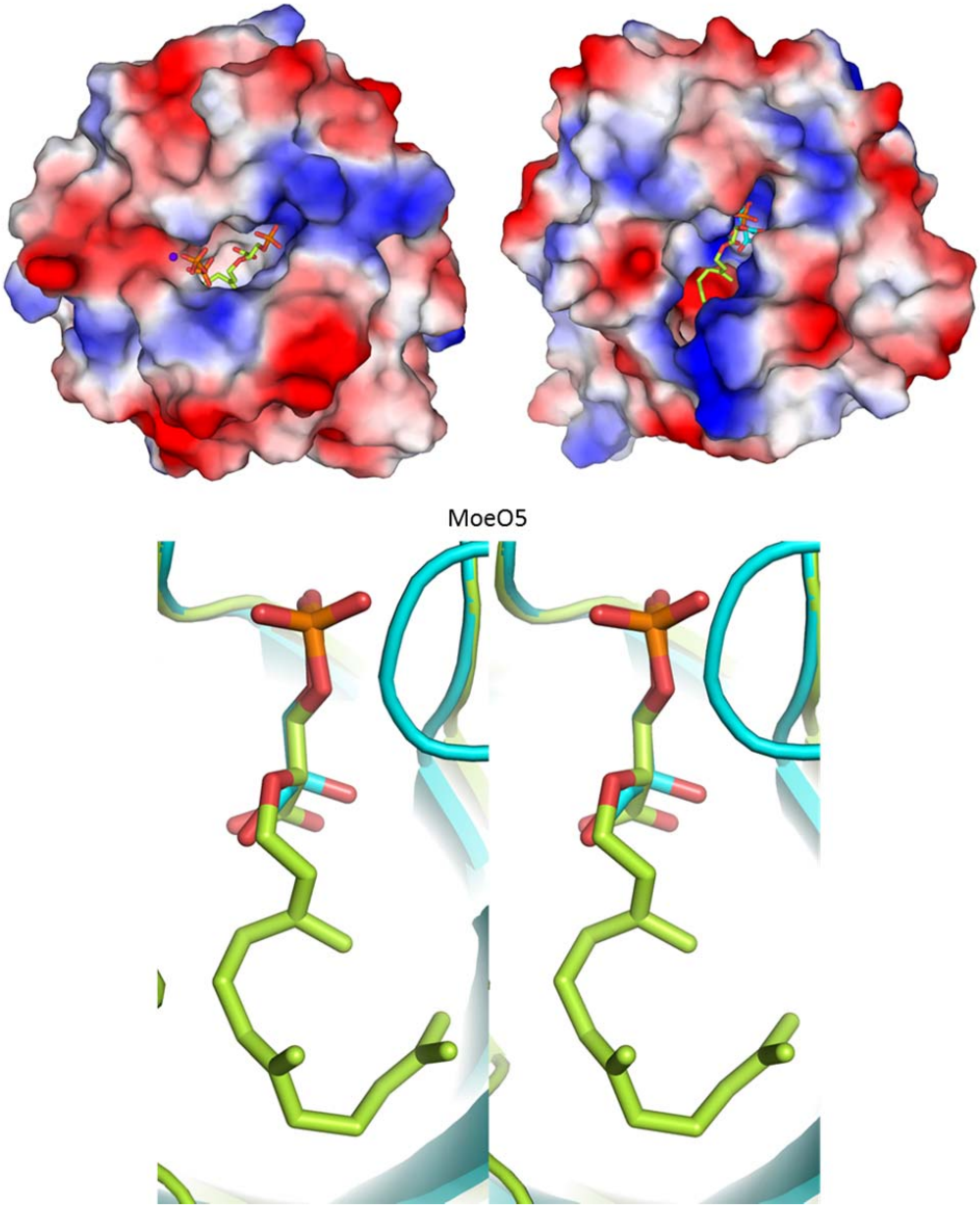


Figure 2.10 MoeO5 interactions with the bound ligands. a) In a stereo view, the FPG is depicted by thick stick model with gray carbon atoms, and the surrounding amino acids are shown as thin sticks with green carbons. Hydrogen bonds between FPG and the protein are shown as dashed lines. The water molecule bound to His97 is consistently observed in all wild-type crystals. b) The bound FsPP and Mg-PPi along with the active-site amino acids are shown in a similar way as in a). The coordinate bonds to the Mg^{2+} ion and other hydrogen bonds to the ligands are shown as dashed lines. Note that PPi is closer to C1 than C3 of the FsPP.

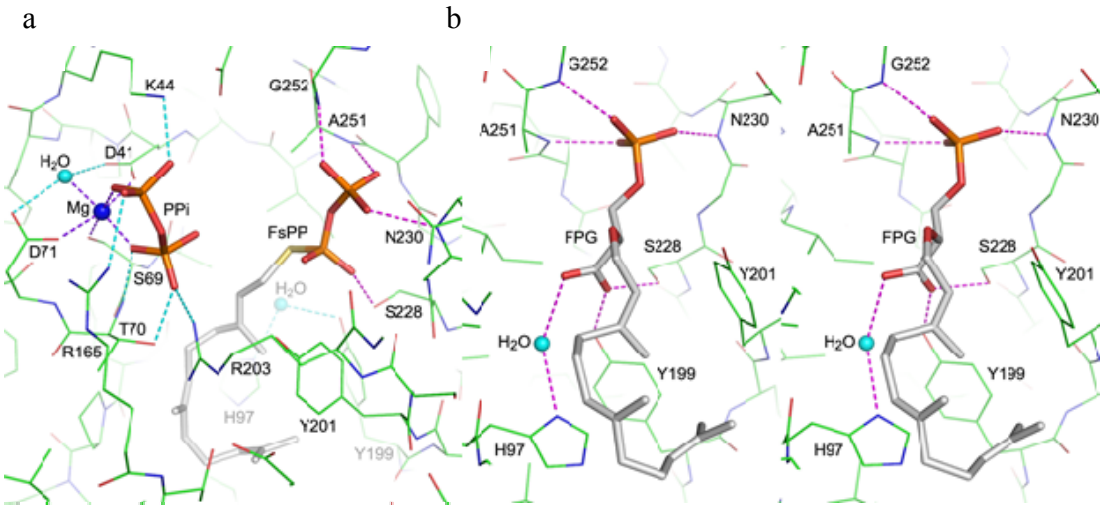
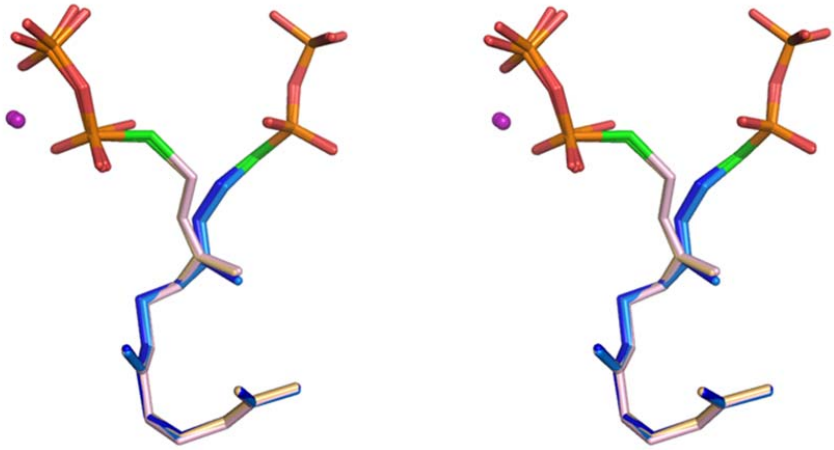
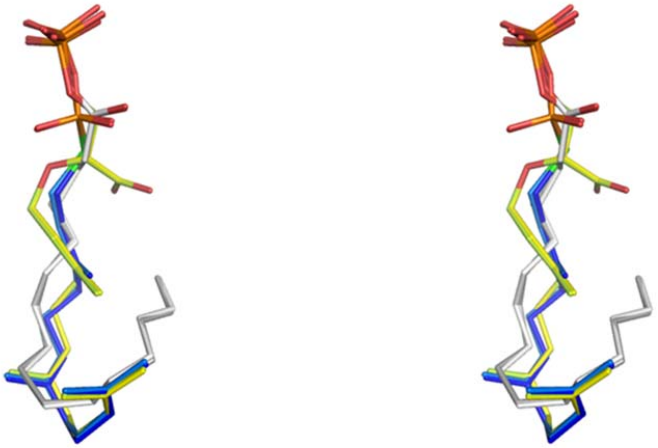


Figure 2.11 Alternative interpretations for the bound ligands. a) The PPi in the bound Mg-PPi was switched with the PPi moiety of FsPP in the model of the overnight soak crystal. The switched model is shown with pink carbon atoms, while the original model has blue carbons. The Mg^{2+} ion is shown as a purple sphere. b) The model for bound ligand in the H97C mutant crystal is shown with carbon atoms colored gray. For comparison, it is superimposed on the FPG (yellow carbons) and FsPP (blue carbons) models.



FsPP (alternative model)



Possible H97C ligand

Figure 2.12 Proposed mechanisms for the TIM-barrel PTs. The reactions start by ionization (bold arrow) of the prenyl pyrophosphate, which requires a bound Mg^{2+}. a) In catalysis by GGGPS and PcrB, the 3'-OH of G1P is deprotonated by Glu167 (Glu160 in PcrB) and associates directly with the carbocation, forming an ether bond. The prenyl chain length varies from n = 3 in GGGPS to n = 5 in PcrB. b) In MoeO5, *trans*-FPP, *Z,E*-FPP and NPP all react to form FPG, consistent with an ionization/isomerization mechanism, but shorter and longer chain species do not. c) His97 is essential for catalysis and may play a role in isomerization, possibly via NOH.

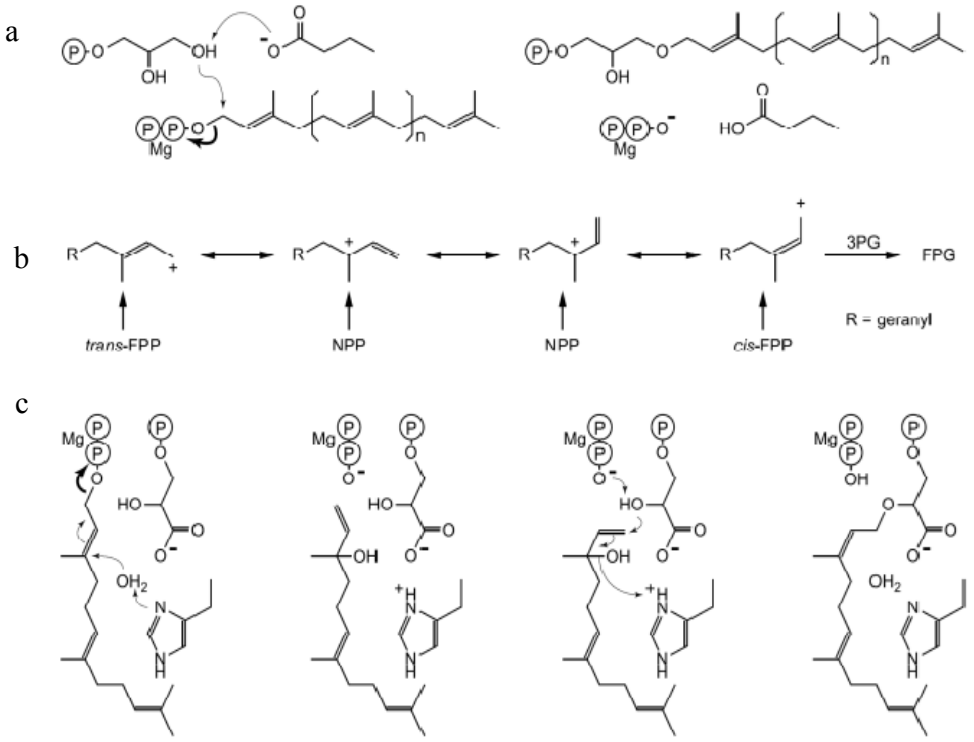
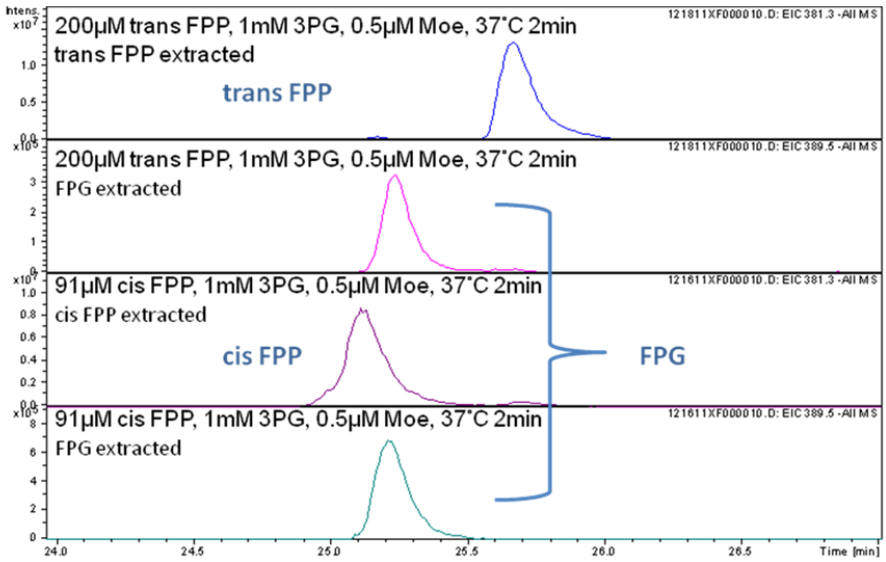


Figure 2.13 *Z,E*-FPP and its reaction rate relative to *trans*-FPP. The upper panel shows HPLC results. Reaction buffer: Tris-HCl, pH 7.5, 1 mM $MgCl_2$; column: C8. The lower panel shows the curves used for calculating the reaction rates. The results indicate V_{max} (*trans*-FPP) = 2.16 μM/min and V_{max} (*cis*-FPP) = 10.63 μM/min, or K_{cat} (*trans*-FPP) = 4.32 min^{-1} and K_{cat} (*cis*-FPP) = 21.26 min^{-1}. That is, *cis*-FPP as substate is ~ 5x more active than is *trans*-FPP, which itself has the same activity as NPP.



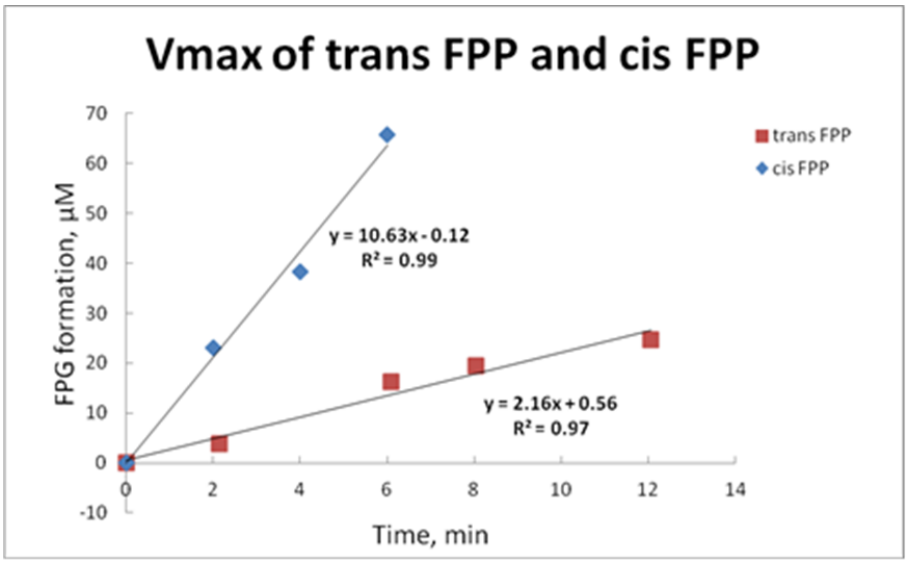
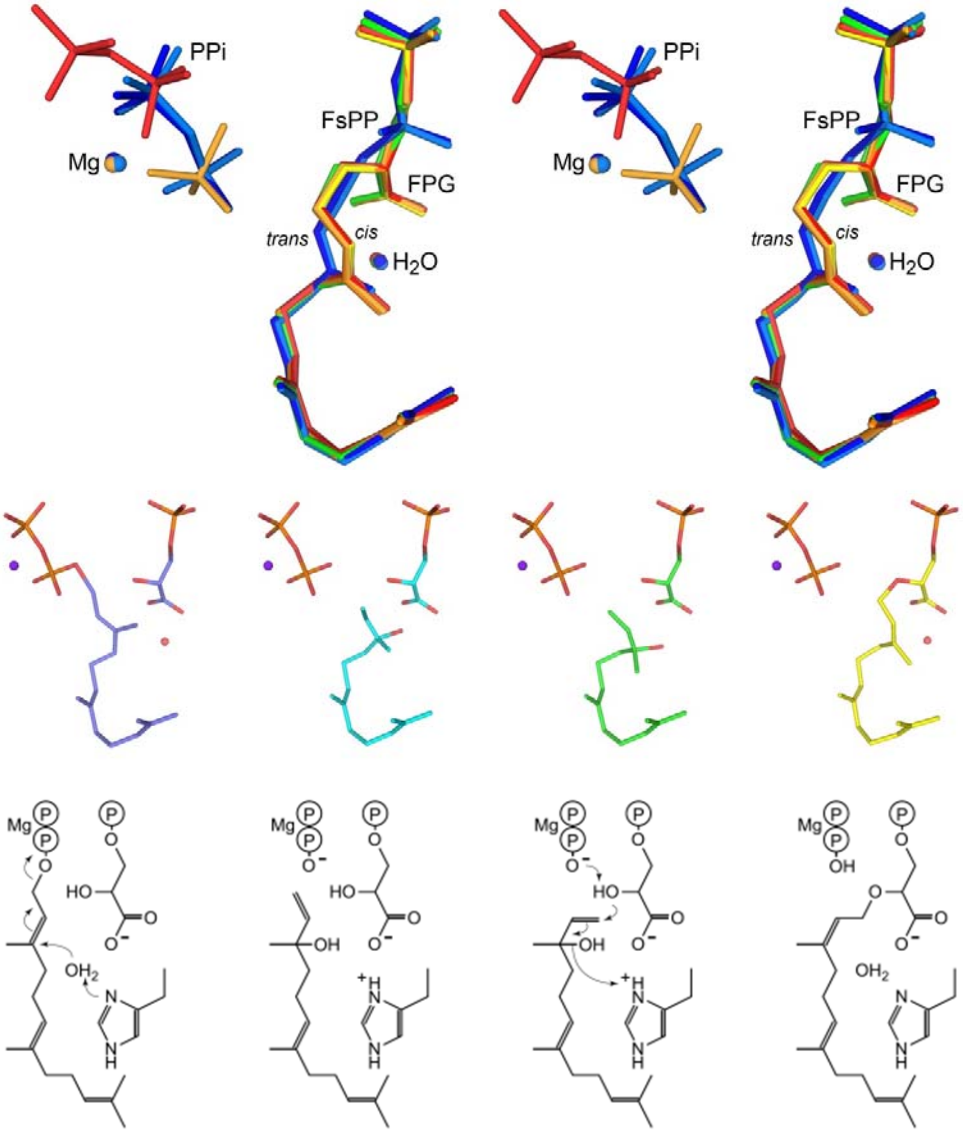


Figure 2.14 Possible catalytic mechanism of MoeO5. The various ligands bound to the wild-type MoeO5 are superimposed and displayed in stereo as ball and stick models. These are colored yellow, green, blue, orange and red for the "native" crystal (though soaked with GGsPP), and crystals soaked with FsPP (3 hours), FsPP (overnight), G3P, and PPi, respectively. The water molecule is 3.4 – 3.6 Å from the C3 atom of the farnesyl group. The PPi is 3.3 Å from the O2' atom of FPG and 3.0 – 3.6 Å from C1, but it is 5.0 – 5.3 Å from C3. A series of models and a proposed catalytic mechanism are followed. Here a vital role is assumed by His97, which activates a water molecule for, and thus allows the C3-C2 bond rotation from *trans*-like to *cis*-like conformation. The negatively charged Mg-PPi then accepts a proton from the 2'-OH group of 3PG, forming the product FPG.



2.7 References

[1] B. Ostash, S. Walker, *Nat. Prod. Rep.* **2010**, *27*, 1594 – 1617.

[2] a) B. Ostash, A. Saghatelian, S. Walker, *Chem. Biol.* **2007**, *14*, 257 – 267; b) B. Ostash, E. H. Doud, C. Lin, I. Ostash, D. L. Perlstein, S. Fuse, M. Wolpert, D. Kahne, *Biochemistry* **2009**, *48*, 8830 – 8841.

[3] J. Payandeh, M. Fujihashi, W. Gillon, E. F. Pai, *J. Biol. Chem.* **2006**, *281*, 6070 – 6078.

[4] E. H. Doud, D. L. Perlstein, M. Wolpert, D. E. Cane, S. Walker, *J. Am. Chem. Soc.* **2011**, *133*, 1270 – 1273.

[5] a) H. Guldan, F. M. Matysik, M. Bocola, R. Sterner, P. Babinger, *Angew. Chem. Int. Ed.* **2011**, *50*, 8188 – 8191; b) J. Badger, J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher et al., *Proteins* **2005**, *60*, 787 – 796.

[6] a) R. T. Guo, T. P. Ko, A. P. Chen, C. J. Kuo, A. H. Wang, P. H. Liang, *J. Biol. Chem.* **2005**, *280*, 20762 – 20774; b) P. H. Liang, *Biochemistry* **2009**, *48*, 6562 – 6570.

Additional references

[7] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307 – 326.

[8] Terwilliger TC, *Methods Enzymol.* **2003**, *374*, 22 – 37.

[9] a) Collaborative Computational Project Number 4, *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760 – 763; b) G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, 67, 355 – 367.

[10] S. X. Cohen, M. B. Jelloul, F. Long, A. Vagin, P. Knipscheer, J. Lebbink, T. K. Sixma, V. S. Lamzin, G. N. Murshudov, A. Perrakis, *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64*, 49 – 60.

[11] A. T. Brunger, *Nat. Protoc.* **2007**, *2*, 2728 – 2733.

[12] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486 – 501.

[13] J. S. Richardson, *Adv. Protein Chem.* **1981**, *34*, 167 – 339.

[14] W. Wang, C. Dong, M. McNeil, D. Kaur, S. Mahapatra, D. C. Crick, J. H. Naismith, *J. Mol. Biol.* **2008**, *381*, 129 – 140.

[15] B. Ostash, E. H. Doud, C. Lin, I. Ostash, D. L. Perlstein, S. Fuse, M. Wolpert, D. Kahne, *Biochemistry* **2009**, *48*, 8830 – 8841.

[16] W. S. J. Valdar, *Proteins* **2002**, *43*, 227 – 241.

[17] C. Cole, J. D. Barber, G. J.Barton, *Nucleic Acids Res.* **2008**, *36*, W197 – 201.

Chapter 3: Insights into TIM-barrel Prenyl Transferase Mechanisms: Crystal Structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*

3.1 Notes and Acknowledgements

F. Ren, X. Feng C.-H. Huang, Y. Hu, H.-C. Chan cloned, expressed and purified PcrB proteins. X. Feng studied product distribution of PcrB and carried out site-directed mutagenesis. F. Ren, T.-P. Ko and R.-T. Guo carried out crystallization, data collection, structure determination and refinement. K. Wang synthesized *trans*-farnesyl pyrophosphate and farnesyl thiopyrophosphate. X. Feng, C.-C. Chen, X. Pang, M. He, Y. Li, E. Oldfield, and R.-T. Guo analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by National Basic Research Program of China (grant 2011CB710800 to RTG), Tianjin Municipal Science and Technology Commission (10ZCKFSY06000 to RTG), and the United States Public Health Service (National Institutes of Health grants GM65307 and AI074233). We thank the National Synchrotron Radiation Research Center of Taiwan for beam-time allocation and data-collection assistance, and P. Babinger for providing coordinates.

This chapter was reproduced in part with permission from F. Ren, X. Feng, T.-P. Ko, C.-H. Huang, Y. Hu, H.-C. Chan, Y.-L. Liu, K. Wang, C.-C. Chen, X. Pang, M. He, Y. Li, E.Oldfield, and R.-T. Guo. *ChemBioChem*, Volume 14, Issue 2, pages 195–199, January 21, 2013. F. Ren, X. Feng and T.-P. Ko contributed equally to this work.

3.2 Introduction

Many enzymes involved in isoprenoid (or terpene/terpenoid) biosynthesis are prenyl transferases (PTs) that catalyze condensation reactions between prenyl diphosphates and various acceptors and structurally, most contain one of six main folds: α, β, γ, δ, ε, or ζ[1]. Recently, several new PT structures that contain a triose phosphate isomerase (TIM) barrel fold have been discovered[2], and there is considerable interest in their mechanisms of action[2b, 2e]. The first structure published was that of *Archaeoglobus fulgidus* geranylgeranylglyceryl phosphate synthase (AfGGGPS)[2d], which catalyzes the reaction

between geranylgeranyl diphosphate (GGPP) and glycerol-1-phosphate (G1P) to form geranylgeranylglyceryl phosphate (GGGP) (Scheme 3.1), a precursor of the (perhydro)geranylgeranyl-based phospholipids found in Archaea[3]. A very similar reaction is carried out by the enzyme PcrB in several Bacilli but in this case the prenyl donor is heptaprenyl diphosphate (HepPP)[4]. The PcrB product, heptaprenylglyceryl phosphate (HepGP), is then dephosphorylated to heptaprenylglycerol, then acetylated, Scheme 3.1[4]. The function of this ester product is unknown. However, PcrB forms part of an operon with PcrA, an essential DNA helicase[5], and a Δ*pcrB* mutant (in *Bacillus subtilis*) has a deranged, "cloggy" pattern of growth[4], so it seems likely that PcrB is important for growth (or perhaps, virulence) in other Bacilli, such as the pathogen *Staphylococcus aureus*, and may be of interest as a drug target.

In GGGPS or PcrB, there is no isomerization of the isoprenoid chain during catalysis. However, in a third TIM-barrel PT, *Streptomyces ghanaensis* MoeO5 (SgMoeO5), involved in the first committed step in biosynthesis of the antibiotic moenomycin[2c], there is a remarkable *trans* → *cis* isomerization during the prenylation reaction, as shown in Scheme 3.1[2b, 2c], and we recently reported the structure of MoeO5 containing its bound product, FPG (2-*(Z, E)*-farnesyl-3-phosphoglycerate)[2e]. Here, we report the structures of PcrB from *B. subtilis* and *S. aureus* containing bound substrates as well as other ligand molecules which help define the catalytic mechanism, in addition to site-directed mutagenesis and bioinformatics results that point to the specific roles of several residues in PcrB catalysis.

3.3 Results and Discussions

BsPcrB structures were solved by using a previous BsPcrB structure (1VIZ)[2a] as a search model, then the new BsPcrB structure was used as a search model for solving the SaPcrB structures by using the molecular replacement method (Table 3.1 and 3.2 in the Supporting Information). Detailed experimental procedures for protein expression, crystallization, data collection and structure determination are provided in the Supporting Information. The electron densities of the ligands were clear and are shown in Figure 3.1 in the Supporting Information. BsPcrB and SaPcrB both crystallize as dimers, and the dimer interface involves mainly hydrophobic residues in helices α4 and α5. This interface

is different from that in the structure (1VIZ) reported previously[2a], but is the same as predicted by Peterhoff et al.[6] (Figure 3.2 in the Supporting Information).

We show in Figure 3.3a the superimposed structures of the monomer subunits of BsPcrB (in green) and SaPcrB (in cyan) and in Figures 3.3b and 3.3c we show the BsPcrB structure (in green) superimposed on that of AfGGGPS (in yellow; Figure 3.3b) and SgMoeO5 (in orange; Figure 3.3c) in which it can be seen that there is high similarity between the structures of the TIM barrel folds in each system with Cα root mean square deviations (RMSD) in the range 0.98 – 1.51 Å over ~210 atoms, Figure 3.3. (The RMSD between BsPcrB and SaPcrB, GGGPS and MoeO5 are 0.98 Å, 1.06 Å and 1.51 Å, respectively, for 210, 204 and 167 matched atoms.)[7] The major difference occurs in the region around loop λ3, which replaces helix α3 in a canonical TIM barrel structure. This loop is rather open in PcrB and GGGPS, acting as a swinging gate to the hydrophobic substrate-binding tunnel, but it collapses inward in MoeO5, making a unique pocket for binding substrate in a bent conformation. [2]

With BsPcrB we also obtained structures of the protein with bound G1P, FsPP (S-*thiolo*-farnesyl diphosphate) and FsPP + G1P (Figure 3.4a), and with SaPcrB we obtained structures with a bound PEG (Figure 3.4b) and a sulfate ion. In BsPcrB-G1P/FsPP the G1P orientation is slightly different from that in the BsPcrB-G1P complex, suggesting flexibility. Consequently, BsPcrB-G1P and BsPcrB-FsPP are used for structural analyses of these two ligands. Detailed interactions of G1P and FsPP are shown in Figures 3.4c and 1d, respectively. The phosphate group of G1P is bound between two loops and forms hydrogen bonds to several peptide NH groups, as well as the side chain of N209. Both the 2-OH and 3-OH are hydrogen bonded to the side chain of E160 and Y158, Figure 3.4c. In the SaPcrB-SO_4 structure, a bound sulfate ion binds to the G1P binding site (Figure 3.5). In the BsPcrB-FsPP structure, the diphosphate group of FsPP is buried in a polar pocket with its prenyl moiety in a straight hydrophobic tunnel, and in the SaPcrB-PEG structure, the PEG polymer is seen in the same hydrophobic tunnel (Figure 3.4d). These structures are of interest since they show the positions of the reactive "head-groups" that condense to form the HepGP product. In addition, they help define experimentally the hydrophobic tunnel into which the C_{35} side-chain can bind, in addition to confirming the nature of the dimer structure predicted by Peterhoff et al[6].

The reaction catalyzed by PcrB involves the attack of the nucleophile G1P via its C-3 OH group at C1 in HepPP. This reaction can reasonably be thought to require: 1) activation of the C-3 OH to facilitate nucleophilic attack; 2) removal of the product diphosphate and 3) a route to removal of the C-3 OH proton that is lost during the reaction, from the active site region. To see which residues are most essential for catalysis and might thus be involved in these steps, we used the SCORECONS program[8] which enables a rank-ordering of residues in terms of their conserved nature.

Many PTs such as farnesyl diphosphate synthase, undecaprenyl diphosphate synthase and squalene synthase, have an absolute requirement for a divalent cation, typically Mg^{2+}[9], and in these systems Mg^{2+} is "anchored" by one or more Asp residues. We thus assumed that there would be a "most essential" Asp in PcrB. However, as can be seen in Table 3.3, the results of the SCORECONS analysis (using JPRED3 [10] alignments based on BsPcrB or SaPcrB) clearly indicate that the three most conserved residues in both species are Y and E: Y118, Y159 and E161 in *S. aureus* and Y118, Y158 and E160 in *B. subtilis*, Table 3.3. On inspection of the crystal structures of BsPcrB (+FsPP+G1P) it can clearly be seen that E160 is only ~2.7Å from the C-3 OH oxygen (Figure 3.6), which could facilitate removal of the OH proton, and two Tyr OH oxygens are also quite close by (both 4.2Å to the E160 carboxyl oxygens) and are poised to aid proton abstraction from the active site. We thus propose that these three residues are involved in C-3 OH activation and H^+ removal from the active site region. We also see, Table 3.3, that D14 is quite highly conserved, and based on an alignment with the MoeO5 structure (which contains an Mg^{2+}, Figure 3.6) we propose that D14 binds to Mg^{2+} which – as with other PTs, is important in removal of the diphosphate product of the reaction. We tested the activity of wild type protein in the presence of 50 μM EDTA as well as in the absence of $MgCl_2$. No product was found, indicating that Mg^{2+} is likely essential for catalysis. We also tested the activity of D14A mutants of both SaPcrB and BsPcrB and found no activity.

The reaction catalyzed by GGGPS is very similar to that catalyzed by PcrB - the only obvious chemical difference being that GGGPS uses a shorter prenyl diphosphate, and a SCORECONS analysis shows that Y124, Y165 and E167 are the three most essential residues and that D13 is also highly conserved, Table 3.3. Other highly conserved

residues are all Glys that are presumably of structural importance. The role of the conserved D76 (or 77) is unknown. Based on these results then, the YYED set of residues is common to PcrB as well as GGGPS proteins and is involved in proton abstraction (YYE) and diphosphate removal (D).

Next, we consider the nature of the hydrophobic pocket that accommodates the C_{35} side-chain. In recent work, Guldan et al.[4] have used computational docking (together with geometry optimization) of HepPP to a BsPcrB dimer model to propose a model for bound HepPP[4]. This model (in yellow) is shown in Figure 3.7a superimposed on the FsPP substrate (in cyan) plus the PEG ligand (in green) and as can be seen in Figure 3.7a, there is very good accord between the disposition of HepPP in the computational model and the positions of the two ligands in the x-ray structures, albeit with some small differences in the head-group region, Figure 3.7a. Also of interest are the positions of residue A100 and Y104: A100 allows binding of substrates longer than GGPP (length defined by FsPP plus PEG), and Y104 locates near the terminus of HepPP and may play a role in chain length limitation. The residue corresponding to A100 in PcrB is W99 in GGGPS, and a previous study on a W99A mutant of GGGPS showed that mutation of this bulky residue to alanine allowed the protein to use substrates longer than GGPP[4]. Interestingly, although Y104 is not highly conserved (Table 3.3), this position is either Y or F in essentially all sequences annotated as HepGPS, as highlighted in green, but is random in the GGGPS sequences. Correspondingly, W100 is conserved in the same GGGPS sequences (highlighted in yellow), but is Ala in the corresponding HepGPS sequences (highlighted in blue), enabling formation in long chain products.

To see to what extent there might be chain length discrimination in PcrB, we produced A100W and Y104A mutants of SaPcrB and tested the effects of these mutations on the conversion of C_{20}, C_{25}, C_{30}, C_{35} and C_{40} prenyl diphosphates to the corresponding glyceryl phosphate products. To investigate these effects, we used a radio-TLC assay with a distribution of polyprenyl diphosphate chain length substrates, obtained from the reaction of FPP with [^{14}C]-IPP, catalyzed by *E.coli* octaprenyl diphosphate synthase[11]. The results obtained, illustrated in Figure 3.7b and Figure 3.8 in the Supporting Information, show that conversion of A100 to W100 results in strong inhibition of the formation of the longer chain products, presumably because the bulky

W100 side-chain inhibits binding of the longer chain substrates, while conversion of Y104 to A104 favors formation of longer-chain products. The observation that the Y104A mutant makes more C_{40} products than does the wild-type protein implies, then, that Y104 is located at or near the terminus of the C_{35} chain since, if it were not, there would be little effect on the chain length distribution. Of course, in cells, the actual product chain length distribution depends on the amounts of FPP, GGPP, HepPP and other prenyl diphosphates that are present, and this is likely to vary with growth conditions. *S. aureus* does, however, make quinones with C_{40} side-chains[12] and it is possible that Y104 contributes to chain-length discrimination in the PcrB reaction since primarily C_{35} products are found in the heptaprenylglycerol diester product[4], but of course chain length discrimination can also occur in reactions occurring after the PcrB-catalyzed reaction.

Finally, we consider the structural differences between the BsPcrB and MoeO5 structures. As can be seen in Figure 3.9, in MoeO5 the prenyl side-chain in the FPG product (in yellow) is highly bent. The chain terminus in this bent structure would occupy the position of the Y118 side-chain in BsPcrB, and binding in this bent manner is only possible because the residue at this position in MoeO5 is an alanine, as shown in Figure 3.9. In addition, the conserved catalytic E160 (in BsPcrB) in MoeO5 is a tyrosine, Y201, which as with BsPcrB, is likely to be involved in H^+ abstraction, this time from the 2-OH group of 3-phosphoglycerate, illustrated in Figure 3.9. The residue corresponding to D14 (BsPcrB) or D13 (GGPPS) in MoeO5 is D41, which interacts with Mg^{2+} in the active site[2e].

3.4 Conclusions

In summary: we report the structures of PcrBs from *B. subtilis* (4 structures) and *S. aureus* (2 structures), the first substrate (or other ligand) containing structure of these enzymes and the first structure of the *S. aureus* protein. The results clearly show the G1P and F(s)PP binding sites, plus, they map out much of the C_{35} side-chain binding pocket, and there is good accord with earlier computational predictions as to the nature of this binding site, as well as the nature of the dimer structure. The most highly conserved residues are two Tyr and one Glu in the active site, proposed to be involved in 3-OH

activation as well as H^+ removal from the active site region, together with a conserved Asp, involved in Mg^{2+} binding and diphosphate removal. In addition, we propose that Y104 is involved in chain-length regulation. The overall structures are similar to those previously reported for MoeO5 but there, the bent FPG side-chain occupies the Y118 (BsPcrB)→ A124 (MoeO5) site, and E160 (BsPcrB) is replaced by Y201 (MoeO5), involved in 2-OH proton abstraction.

3.5 Materials and Methods

Expression and purification of PcrB from *B. subtilis* and *S. aureus*

The genes encoding PcrB from *Bacillus subtilis subsp. subtilis str. 168* and *Staphylococcus aureus subsp. aureus Mu3* were amplified by polymerase chain reaction (PCR) with a forward primer 5'-GGTATTGAGGGTCGCGCTGGTGCTGGTGCTATGT ACGATGTAACGGAGTGGAAGCAT-3' and a reverse primer 5'-AGAGGAGAGTTAG AGCCATTACTCGCCTTTCACAGCCGCTACTGT-3' for BsPcrB; and a forward primer 5'-GGTATTGAGGGTCGCGCTGGTGCTGGTGCTATGTATGACATTAAAAA ATGGCGCCAT-3' and a reverse primer 5'-AGAGGAGAGTTAGAGCCATTATTTACT AGACTCCTTTATTTTTAC-3' for SaPcrB, and then cloned into the pET32 Xa/LIC vector. Mutants were prepared with BsPcrB-pET32a Xa/LIC and SaPcrB-pET32a Xa/LIC plasmids as the templates. The recombinant plasmids were transformed to *E. coli* BL21 (DE3) and proteins were induced with 1 mM isopropyl-thiogalactopyranoside (IPTG) at 30 °C for 12 hours. The cell pastes were harvested by centrifugation at 7,000 x *g* and resuspended in a lysis buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM NaCl and 20 mM imidazole. Cell lysates were prepared with a French Press Instrument, and then centrifuged at 17,000 x *g* to remove cell debris. Proteins were purified by FPLC using a Ni-NTA column. The buffer and gradient for the Ni-NTA column were 25 mM Tris, pH 7.5, 150 mM NaCl, and 20–250 mM imidazole. The proteins were eluted at about 130 mM imidazole for BsPcrB and 135 mM imidazole for SaPcrB. The proteins were dialyzed against the buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM NaCl and then subjected to Factor Xa digestion to remove the tag. The mixture was then passed through another Ni-NTA column and subsequently untagged BsPcrB and SaPcrB were eluted with 10 mM imidazole-containing buffer and then dialyzed twice against 5 L

buffer (25 mM Tris-HCl, pH 7.5, and 150 mM NaCl) for storage. SDS-PAGE analysis was used to check the purity (>95%) of GGPPs and its mutants.

Crystallization, data collection, structure determination and refinement

Crystals of apo BsPcrB were prepared by mixing 2 μL protein solution (10 mg/mL in 25 mM Tris-HCl, pH 7.5 and 10 mM $MgCl_2$) with equal amounts of mixture solution and mother liquor, and equilibrating with 500 μL of the mother liquor at room temperature by using the sitting drop method using Hampton Research (Laguna Niguel, CA) kits. The monoclinic BsPcrB crystal was obtained from a solution that contained 2.5 M NaCl and 14% PEG6000. The BsPcrB-FsPP crystal is obtained by co-crystallizing BsPcrB protein solution with 5 mM FsPP under the same conditions. The BsPcrB-G1P crystal was prepared by soaking the apo BsPcrB crystals with 50 mM G1P in cryoprotectant (2.5 M sodium chloride, 17% PEG 6000, 10 mM $MgCl_2$, and 5% glycerol) for 3 h. The BsPcrB-G1P/FsPP crystals were obtained by soaking the BsPcrB-FsPP crystals with 50 mM G1P and 5 mM FsPP in the same cryoprotectant mentioned above.

Crystals of the SaPcrB were prepared by mixing 2 μL protein solution (10 mg/ml in 25 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$) with equal amounts of mixture solution and mother liquor, and equilibrating with 500 μL of the mother liquor at room temperature by using the sitting drop method. Two SaPcrB structures with different space groups were solved. Orthorhombic SaPcrB crystals ($P2_12_12_1$) were obtained from a solution that contained 0.1 M Tris, pH 8.5, and 21% PEG3350, after 3 days. Tetragonal SaPcrB crystals ($P4_12_12$) were obtained from 0.1 M Tris, pH 8.5, and 2.5 M ammonium sulfate after 2 days.

Diffraction data were collected at beam lines BL13B1 and BL13C1 of the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan), and processed using the program HKL2000[13]. Prior to structural refinements, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor.[14] The BsPcrB complex structures solved in this study were determined by using the previously solved BsPcrB structure (1VIZ). The initial phase of SaPcrB was determined by using the molecular replacement method (MR) with BsPcrB-FsPP as the search model. All following structural refinements were carried out using the programs Coot[15] and CNS[16]. Data

collection and refinement statistics for BsPcrB and SaPcrB crystals are summarized in Table 3.1 and 3.2. All diagrams of the protein structures were prepared by using PyMOL (http://www.pymol.org/).

Kinetic measurements on SaPcrB proteins

Expression and purification of *E. coli* OPPS were carried out as described previously[11]. ^{14}C labeled prenyl diphosphates with C_{20} to C_{50} were synthesized by adding 1 mM FPP, 20 μM GGPP, 100 μM [^{14}C]IPP and 1 μM EcOPPS into 50 μL of 50 mM HEPES (pH 7.5), 50 mM KCl and 0.5 mM $MgCl_2$, and reacting at 37°C for 40 min. Equal amount of SaPcrB wild type or mutants, together with 4 mM $MgCl_2$ and 4 mM glycerol phosphate, were added to the mixture of ^{14}C labeled prenyl diphosphate substrates, and the resulting reaction mixture incubated at 37°C for 6 hours. Dephosphorylation of the PcrB product (polyprenyl glycerol phosphate) was then carried out by adding potato acid phosphatase and 100 μL of 40 mM sodium acetate (pH 4.7), 20% 1-propanol and 0.1% Triton X-100. The reactions were then terminated after overnight incubation at 37°C by adding 500 μL NaCl saturated H_2O, and the polyprenyl glycerol extracted with n-hexane. The solution was then concentrated and spotted onto a C18 reverse phase TLC plate. The TLC plate was developed with acetone/methanol (2:8) and the product distribution analyzed by using a Storm 840 PhosphorImager and Quantity One (Bio-Rad, Hercules, CA).

3.6 Schemes, Charts, Tables and Figures

Scheme 3.1 Structures of molecules and reactions of interest. a) Reactions catalyzed by PcrB (n=5) or GGGPS (n=2) b) Reaction catalyzed by MoeO5.

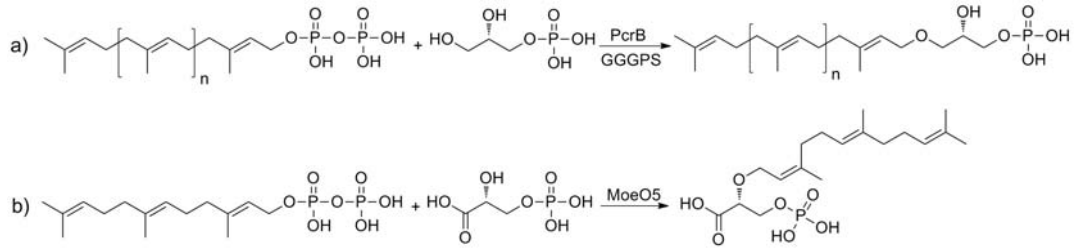


Table 3.1. Data collection and refinement statistics for BsPcrB crystals

	BsPcrB-apo	BsPcrB-G1P	BsPcrB-FsPP	BsPcrB-G1P/FsPP
PDB code	3VZX	3VZY	3VZZ	3W00
Data collection				
space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
unit-cell				
a [Å]	54.1	53.9	54.1	54.1
b [Å]	56.0	55.8	56.1	56.4
c [Å]	77.3	76.7	77.5	77.7
β [°]	108.5	108.8	108.0	108.1
resolution [Å] [a]	25-1.54 (1.60-1.54)	50-1.63 (1.69-1.63)	25-2.04 (2.11-2.04)	25-2.50 (2.59-2.50)
unique reflections	63317 (6150)	52267 (4962)	27153 (2560)	15342 (1440)
redundancy	3.9 (3.3)	3.7 (3.6)	4.0 (3.0)	3.4 (2.2)
completeness [%]	97.9 (95.5)	96.9 (92.9)	96.4 (91.7)	98.6 (92.8)
average $I/\sigma(I)$	25.7 (3.0)	23.3 (3.7)	21.7 (3.4)	16.1 (2.7)
R_{merge} [%]	5.3 (44.2)	4.9 (28.5)	5.9 (31.9)	9.2 (31.0)
Refinement				
no. of reflections	61084 (4969)	51321 (4368)	24893 (1289)	13598 (779)
R_{work} (95% of data)	0.182 (0.248)	0.175 (0.231)	0.178 (0.213)	0.187 (0.221)
R_{free} (5 % of data)	0.225 (0.283)	0.215 (0.264)	0.246 (0.265)	0.259 (0.338)
r.m.s.d. bonds [Å]	0.020	0.020	0.020	0.020
r.m.s.d. angles [°]	2.0	2.0	2.0	2.1
dihedral angles				
most favored [%]	91.7	90.4	91.2	89.2
allowed [%]	7.6	9.0	8.1	10.6
disallowed [%]	0.7	0.5	0.7	0.2
no. of non-H atoms / average B [$Å^2$]				
protein	3508 / 21.5	3504 /22.1	3504 / 25.2	3486 / 25.5
water	529 / 34.6	503 / 34.6	308 / 33.9	153 / 29.7
ligand	15 / 25.9	25 / 29.6	28 / 45.1	39 / 61.7

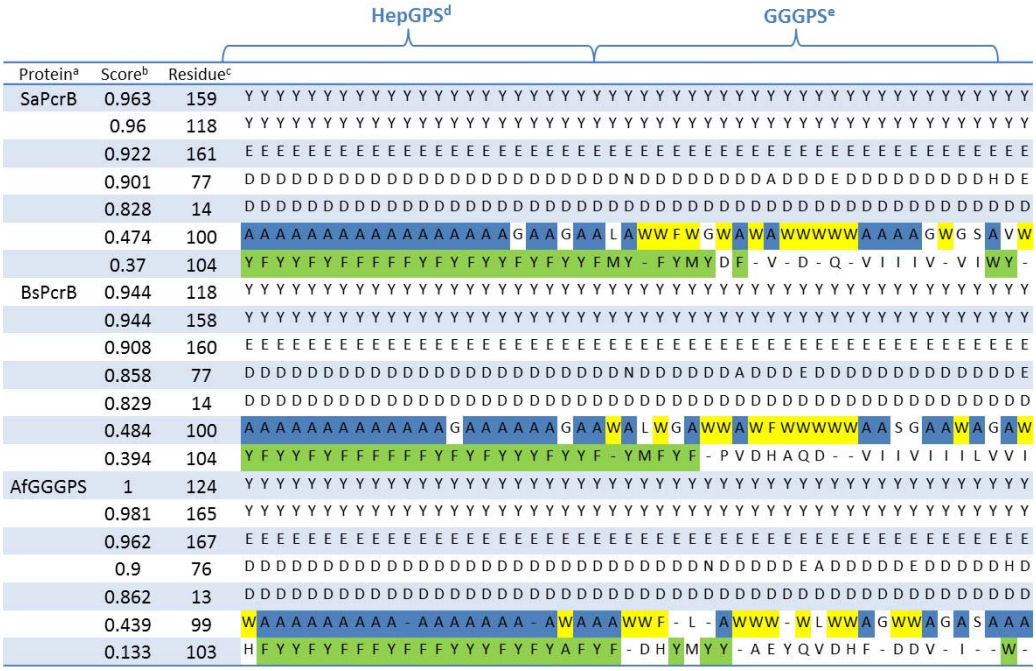
[a]Values in the parentheses are for the highest resolution shells.

Table 3.2 Data collection and refinement statistics for SaPcrB crystals

	SaPcrB-PEG	SaPcrB-SO_4
PDB code	3W01	3W02
Data collection		
space group	$P2_12_12_1$	$P4_12_12$
unit-cell		
a [Å]	65.3	141.6
b [Å]	80.8	141.6
c [Å]	85.8	107.9
β [°]	90.0	90.0
resolution [Å] [a]	25-1.54 (1.60-1.54)	25-2.98 (3.09-2.98)
unique reflections	66696 (5680)	22795 (1987)
redundancy	7.6 (5.2)	6.4 (4.8)
completeness [%]	98.1 (84.5)	98.4 (87.4)
average $I/\sigma(I)$	44.8 (3.7)	31.9 (2.5)
R_{merge} [%]	5.3 (44.9)	5.6 (47.9)
Refinement		
no. of reflections	64963 (4884)	20225 (1050)
R_{work} (95% of data)	0.184 (0.261)	0.194 (0.258)
R_{free} (5 % of data)	0.218 (0.280)	0.234 (0.268)
r.m.s.d. bonds [Å]	0.020	0.020
r.m.s.d. angles [°]	1.9	2.0
dihedral angles		
most favored [%]	93.0	83.3
allowed [%]	6.8	16.6
disallowed [%]	0.2	0
no. of non-H atoms / average B [$Å^2$]		
protein	3584 / 23.4	3573 / 34.8
water	608 / 37.8	94 / 31.0
ligand	10 / 44.7	20 / 70.5

[a]Values in the parentheses are for the highest resolution shells.

Table 3.3 Scorecons results for BsPcrB, SaPcrB and AfGGGPS



[a] Protein used as input to JPRED3[10]; alignment used as input to the SCORECONS[8] program.

[b] Conservation scores from SCORECONS. 1.000 means the most highly conserved residue.

[c] SCORECONS residue number in query protein sequence.

[d] Vertical columns show residues of interest for a given protein. 16/25 proteins shown annotate as HepGPS (=PcrB or PcrB-like).

[e] as d but here 17/25 proteins annotate as GGGPS. The Trp residues (shown in yellow) are involved in chain-length regulation in GGGPS: not found (typically W→A) in HepGPS. Note that residues at the 103/104 positions in GGGPS are essentially random, since they are not involved in chain length regulation.

Figure 3.1 Electron density maps of the bound ligands. The Fo-Fc maps without including any ligand in the refinement steps are contoured at 1.0 and 2.0 σ level, and are colored in green and red. a) G1P in BsPcrB-G1P; b) FsPP in BsPcrB-FsPP; c) G1P and FsPP in BsPcrB-G1P/FsPP; d) PEG in SaPcrB-PEG; e) SO4 ion in SaPcrB-SO4.

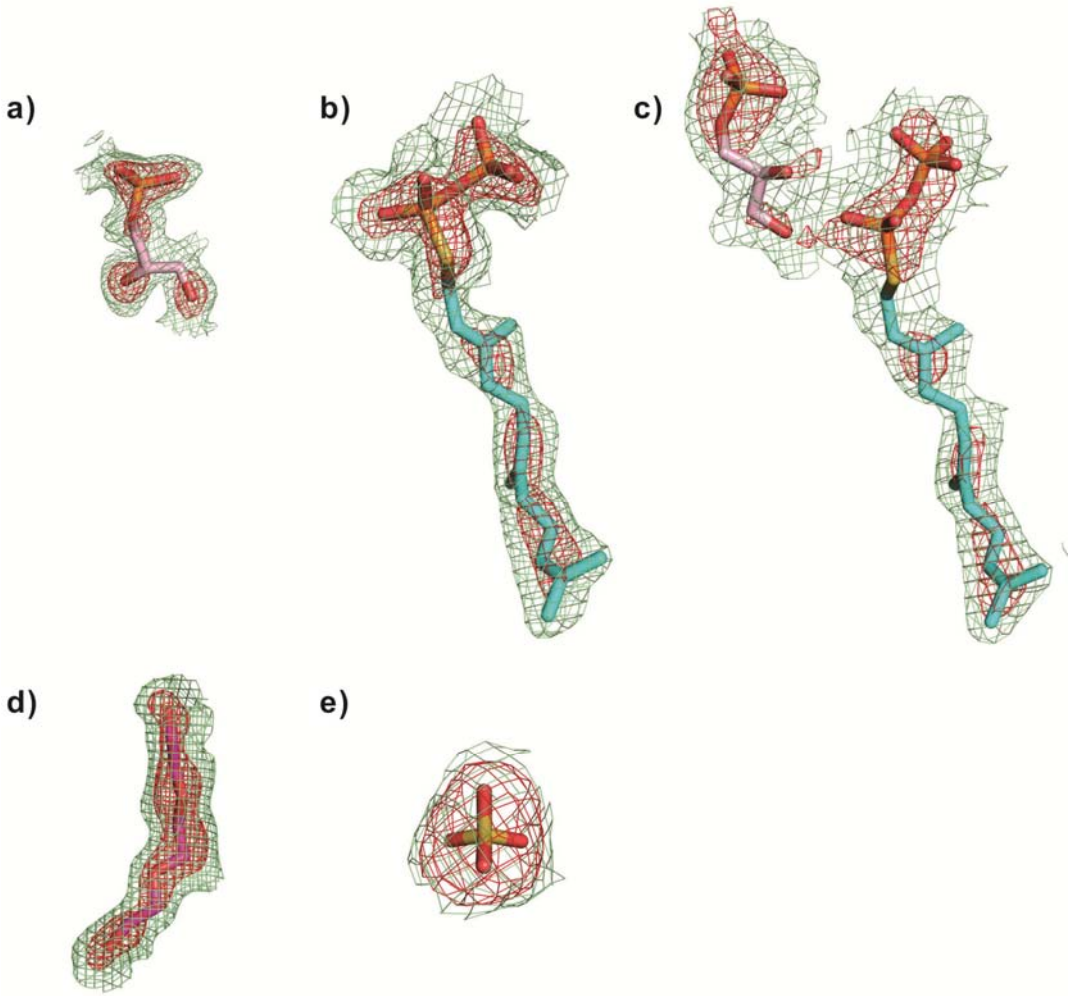


Figure 3.2 Dimer interface of PcrB. a) IVIZ dimer interface; b) Dimer interface of BsPcrB; c) predicted dimer interface proposed by Peterhoff et al[6].

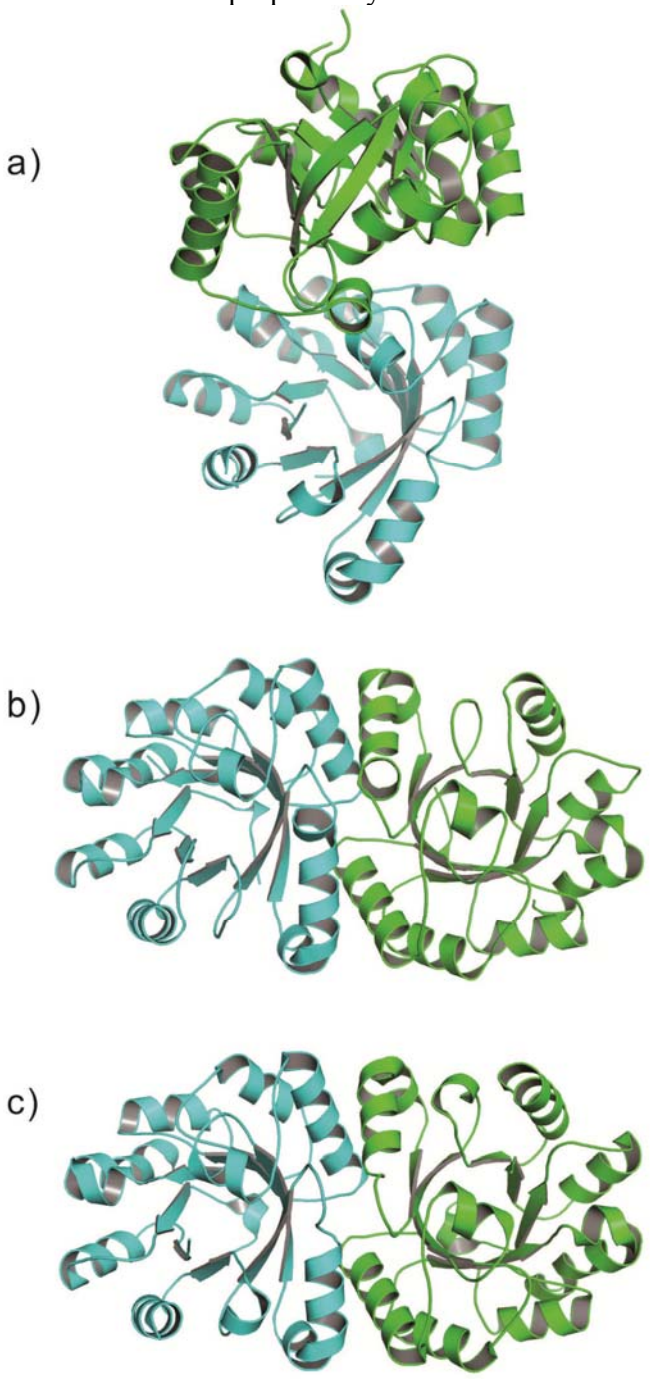


Figure 3.3 Structure comparison of monomers of BsPcrB, SaPcrB, AfGGGPS and SgMoeO5. a) Superimposition of BsPcrB (in green) and SaPcrB (in cyan). b) Superimposition of BsPcrB (in green) and AfGGGS (in yellow). c) Superimposition of BsPcrB (in green) and SgMoeO5 (in orange).

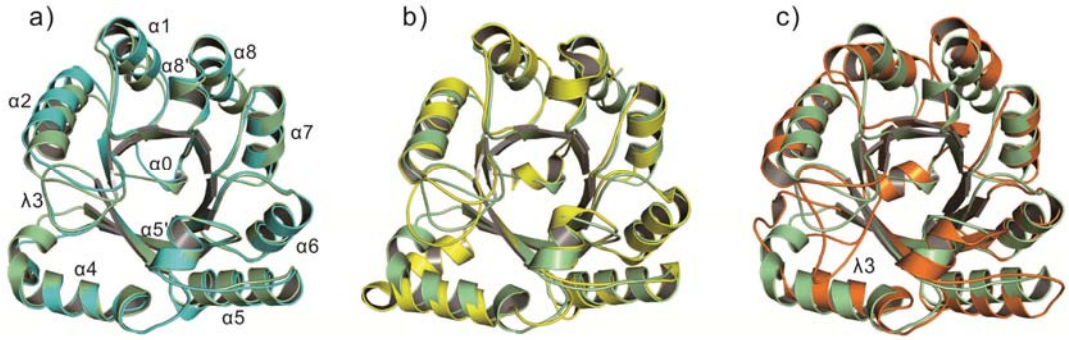


Figure 3.4 PcrB complex structures. a) BsPcrB-FsPP-G1P structure is shown with FsPP and G1P as a stick model and BsPcrB as a cartoon. The ligands in BsPcrB-FsPP and BsPcrB-G1P are very close to those in BsPcrB-FsPP-G1P. b) SaPcrB-PEG structure is shown with PEG as a stick model and SaPcrB as a cartoon. c) Detailed ligand-protein interactions in BsPcrB-G1P. d) Detailed ligand-protein interactions in BsPcrB-FsPP.

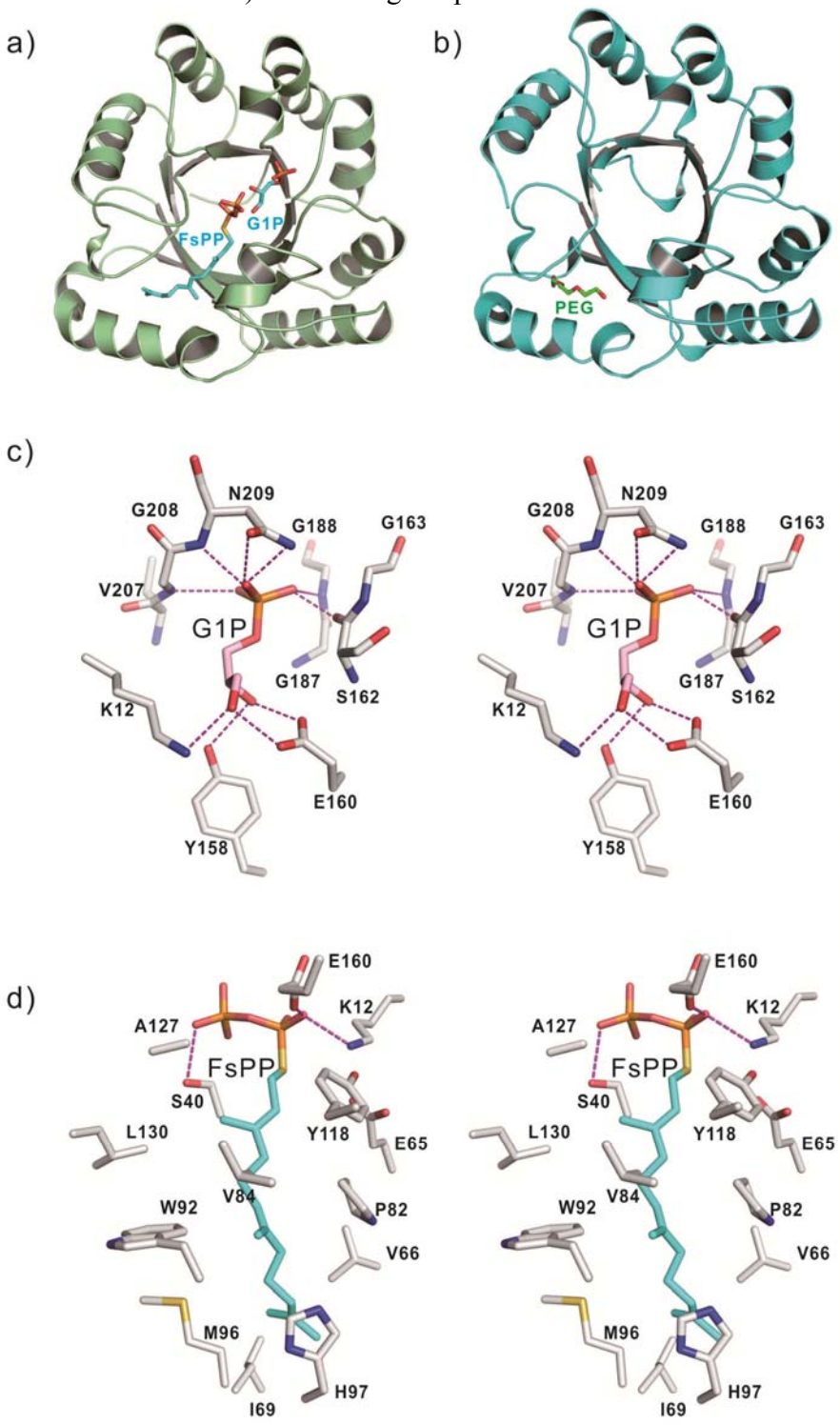


Figure 3.5 Superimposition of SO4 ion and G1P from SaPcrB-SO4 and BsPcrB-G1P/FsPP, respectively. The SO4 ion is colored in black, and the G1P and FsPP are colored in cyan.

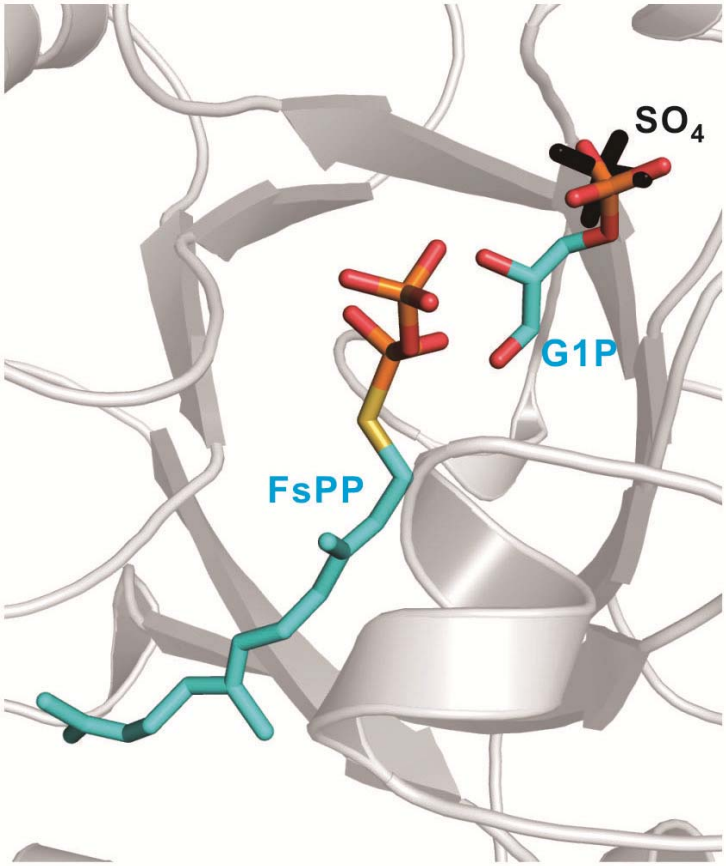


Figure 3.6 Most conserved residues in BsPcrB shown together with FsPP and G1P, as well as a catalytic Mg^{2+} from a superimposed MoeO5 structure.

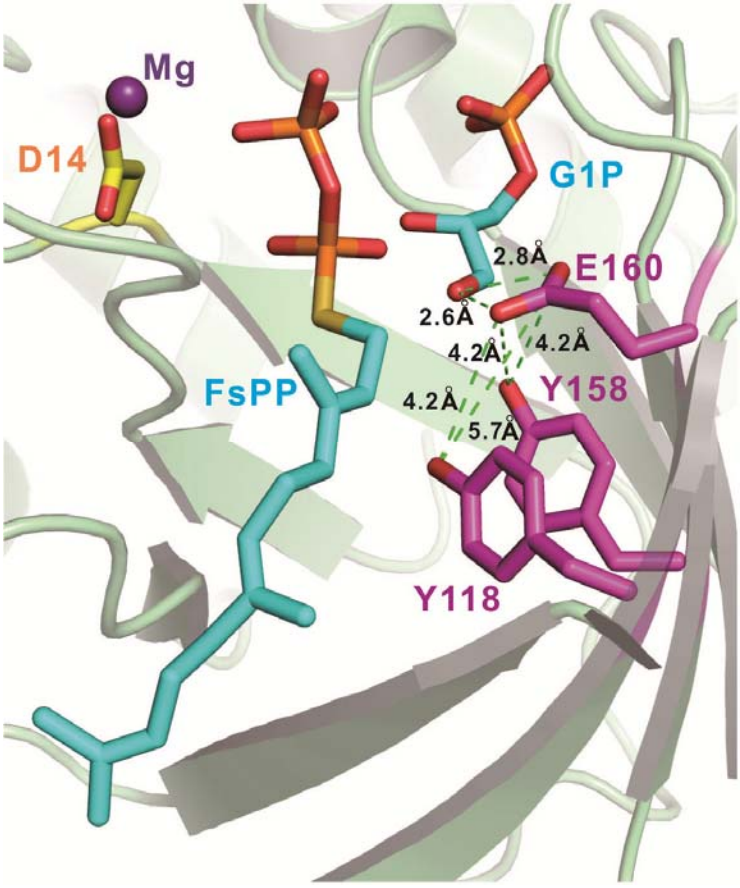


Figure 3.7 Hydrophobic tunnel in BsPcrB with TLC product distribution results. a) HepPP model (in yellow) by Guldan et al. is shown superimposed on FsPP (in cyan) from the BsPcrB-FsPP-G1P structure and PEG (in green) from the SaPcrB-PEG structure. b) TLC product distribution results for SaPcrB mutants with different chain length prenyl diphosphate substrates.

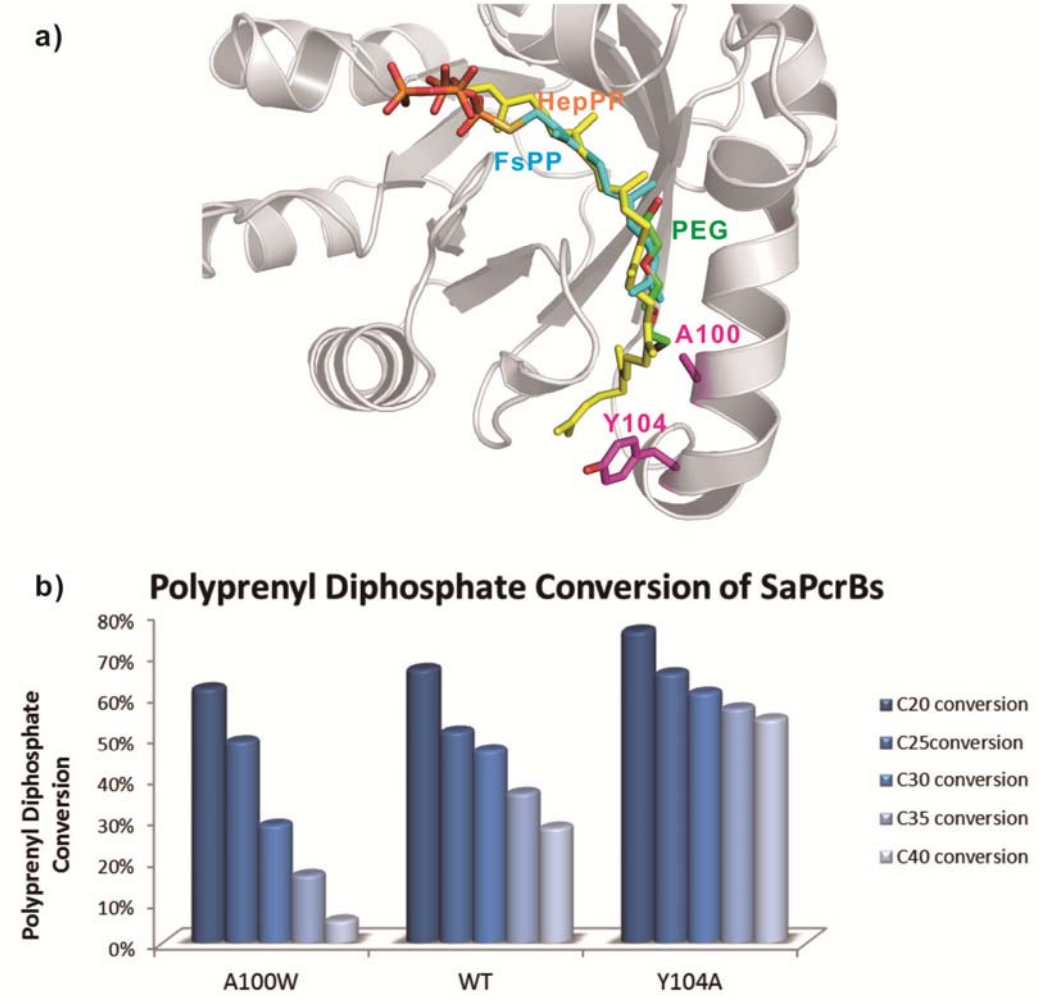
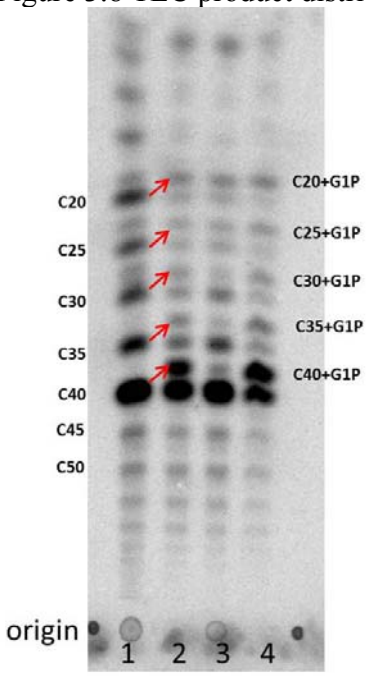


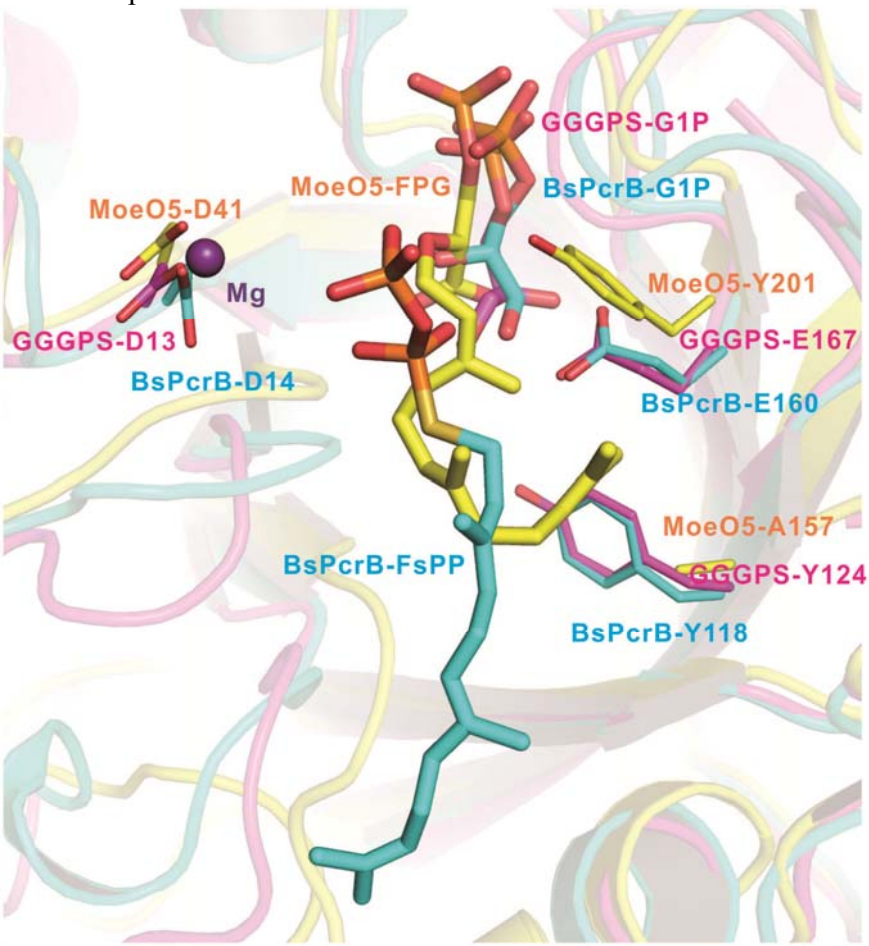
Figure 3.8 TLC product distribution results for SaPcrB mutants.



Lane 1: 1 mM FPP/100 μM [^{14}C]IPP OPPS reaction combined with 20 μM GGPP/100 μM [^{14}C]IPP OPPS reaction;
Lane 2: lane 1 reaction plus SaPcrB_WT;
Lane 3: lane 1 reaction plus SaPcrB_A100W;
Lane 4: lane 1 reaction plus SaPcrB_Y104A;

- Buffer used is 20 mM Hepes (pH 7.5), 50 mM KCl and 4.5 mM $MgCl_2$, total concentration of G1P and G3P is 4 mM.
- Same concentration of PcrBs are used.

Figure 3.9 Structures of BsPcrB ligands (FsPP and G1P, in cyan), GGGPS ligands (G1P, in magenta) and MoeO5 ligands (FGP, in yellow). Conserved residues are also shown as stick models for the proteins.



3.7 References

[1] E. Oldfield, F. Y. Lin, *Angew. Chem.* **2012**, *51*, 1124-1137.

[2] a) J. Badger, J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda, P. Kearins, C. Kissinger, B. Laubert, H. A. Lewis, J. Lin, K. Loomis, D. Lorimer, G. Louie, M. Maletic, C. D. Marsh, I. Miller, J. Molinari, H. J. Muller-Dieckmann, J. M. Newman, B. W. Noland, B. Pagarigan, F. Park, T. S. Peat, K. W. Post, S. Radojicic, A. Ramos, R. Romero, M. E. Rutter, W. E. Sanderson, K. D. Schwinn, J. Tresser, J. Winhoven, T. A. Wright, L. Wu, J. Xu, T. J. R. Harris, *Proteins Struct. Funct. Bioinf.* **2005**, *60*, 787-796; b) E. H. Doud, D. L. Perlstein, M. Wolpert, D. E. Cane, S. Walker, *J. Am. Chem. Soc.* **2011**, *133*, 1270-1273; c) B. Ostash, E. H. Doud, C. Lin, I. Ostash, D. L. Perlstein, S. Fuse, M. Wolpert, D. Kahne, S. Walker, *Biochemistry* **2009**, *48*, 8830-8841; d) J. Payandeh, M. Fujihashi, W. Gillon, E. F. Pai, *J. Biol. Chem.* **2006**, *281*, 6070-6078; e) F. Ren, T.-P. Ko, X. Feng, C.-H. Huang, H.-C. Chan, Y. Hu, K. Wang, Y. Ma, P.-H. Liang, A. H. J. Wang, E. Oldfield, R.-T. Guo, *Angew. Chem.* **2012**, *51*, 4157-4160.

[3] R. Matsumi, H. Atomi, A. J. M. Driessen, J. van der Oost, *Res. Microbiol.* **2011**, *162*, 39-52.

[4] H. Guldan, F.-M. Matysik, M. Bocola, R. Sterner, P. Babinger, *Angew. Chem.* **2011**, *50*, 8188-8191.

[5] M. A. Petit, E. Dervyn, M. Rose, K. D. Entian, S. McGovern, S. D. Ehrlich, C. Bruand, *Mol. Microbiol.* **1998**, *29*, 261-273.

[6] D. Peterhoff, H. Zellner, H. Guldan, R. Merkl, R. Sterner, P. Babinger, *ChemBioChem* **2012**, *13*, 1710-1710.

[7] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126-2132.

[8] W. S. J. Valdar, *Funct. Bioinf.* **2002**, *48*, 227-241.

[9] D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412-3442.

[10] C. Cole, J. D. Barber, G. J. Barton, *Nucleic Acids Res.* **2008**, *36*, W197-W201.

[11] J. J. Pan, T. H. Kuo, Y. K. Chen, L. W. Yang, P. H. Liang, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2002**, *1594*, 64-73.

[12] M. R. Nahaie, M. Goodfellow, D. E. Minnikin, V. Hajek, *J. Gen. Microbiol.* **1984**, *130*, 2427-2437.

[13] z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307-326.

[14] A. T. Brunger, *Acta Crystallogr. Sect. D* **1993**, *49*, 24-36.

[15] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486-501.

[16] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. Sect. D* **1998**, *54*, 905-921.

Chapter 4: Structure and Function of *Bacillus subtilis* YisP: Biofilm Production and Virulence in Bacteria

4.1 Notes and Acknowledgements

Y. Hu, Y. Zheng, C.-H. Huang and X. Feng cloned, expressed and purified BsYisP proteins. Y. Hu, Y. Zheng, H.-C. Chan, T.-P. Ko and R.-T. Guo carried out crystallization, data collection, structure determination and refinement. X. Feng and S. Bogue characterized the enzymatic activity of BsYisP and carried out mutagenesis study. K. Li performed *Bacillus subtilis* pigment extraction experiments. W. Zhu performed DSC experiments. D. López and R. Kolter performed biofilm formation experiments. M. Nega and F. Götz provided staphyloxanthin. C. Nakano and T. Hoshino provided tuberculosinol. X. Feng, E. Oldfield and R.-T. Guo analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by grants from the National High Technology Research and Development Program of China (2012AA022200), the National Basic Research Program of China (2011CBA00805 and 2011CB710800) and the United States Public Health Service (National Institutes of Health grant GM065307). X.F. was supported by a Pre-doctoral Fellowship from the American Heart Association, Midwest Affiliate (13PRE14510056). The synchrotron data collection was conducted on beamline BL13B1 of the NSRRC (National Synchrotron Radiation Research Center, Taiwan) supported by the National Science Council (NSC), Taiwan.

This chapter was reproduced in part with permission from X. Feng, Y. Hu, Y. Zheng, W. Zhu, K. Li, C.-H. Huang, T.-P. Ko, F. Ren, H.-C. Chan, M. Nega, S. Bogue, D. López, R. Kolter, F. Götz, C. Nakano, T. Hoshino, R.-T. Guo, E. Oldfield, paper submitted to *J. Am. Chem. Soc.* on March 10, 2014. X. Feng, Y. Hu and Y. Zheng contributed equally to this work.

4.2 Introduction

There is currently considerable interest in the structure, function and inhibition of the enzymes involved in isoprenoid biosynthesis since in many instances these enzymes represent new drug targets.[1,2] One such class of molecules catalyze the “head-to-head”

condensation of isoprenoid diphosphates to form tri- terpenes such as squalene and dehydrosqualene, reactions catalyzed by squalene synthase (SQS) and dehydrosqualene synthase (CrtM), respectively (Scheme 4.1a). Inhibition of squalene synthase is of interest in the context of the development of cholesterol-lowering drugs[3], and of antiparasitics targeting ergosterol biosynthesis[4].

CrtM is of interest since in *Staphylococcus aureus* it is involved in formation of staphyloxanthin[5,6], a virulence factor for the organism providing resistance to killing from host reactive oxygen species.[7,8] Interestingly, staphyloxanthin levels are elevated in *S. aureus* biofilms[9] and the CrtM inhibitor zaragozic acid inhibits biofilm formation, apparently by affecting lipid micro-domain ("raft") organization.[10] An effect of staphyloxanthin on bacterial membrane structure (order) has also been shown via fluorescence depolarization (of a 1,3,5-diphenylhexatriene probe) and cells with more rigid or ordered membranes appear to be more resistant to host defense peptides as well as to the antibiotic daptomycin.[11]

The CrtM gene (or a close analog) is also present in *Bacillus subtilis* (*yisP*) and it was recently reported[10] that YisP was involved in biofilm and pigment formation and had squalene synthase activity which was blocked by zaragozic acid (Scheme 4.1b), a known SQS/CrtM inhibitor.[12] Zaragozic acid was also reported to block biofilm formation in *S. aureus*, but had no effect on cell growth in either organism. To learn more about the structure and function of YisP, we have solved its structure by x-ray crystallography and determined the structure of the product formed by YisP from farnesyl diphosphate (FPP). The results obtained lead to a proposed mechanism of action model for both this product as well as the CrtM product, staphyloxanthin, on biofilm formation and with staphyloxanthin, on its role as a virulence factor.

With YisP, there are four main questions of interest: What is its structure? What does it make? How does it function? How do its products work? Protein structure predictions of YisP using the Phyre2 program[13] indicate an ε-fold[1] structure akin to that seen in SQS and CrtM, making it likely that the BsYisP product would be a C_{30} isoprenoid. The BsYisP sequence is shown in the alignment in Figure 4.1a and it appears on initial inspection to contain the two expected "Asp-rich" (DDXXD/E and DXXXD) domains[14] found in many prenyl transferases, with DDLDE and a DLGTD motifs being found in the

sequence. However, as can be seen in the alignment of YisP, CrtM and SQS in Figure 4.1a, there are some puzzling features. First, the region of the first aspartate-rich motif (FARM) seen in the alignments of many prenyl synthases, in YisP, is aspartate-poor: AASAD. Second, the DDLDE motif is actually between the FARM and second aspartate-rich motif (SARM) found in HsSQS and CrtM. These results suggest either that YisP might not make squalene (or dehydrosqualene), or that it has an unusual fold that is not suggested by the sequence alignments.

4.3 Results and Discussions

In recent work[15] we cloned, expressed, purified, crystallized and obtained diffraction data on *B. subtilis* YisP. The crystals belonged to the orthorhombic space group $P2_12_12_1$ with unit-cell parameters a = 43.966 Å, b = 77.576 Å, c = 91.378 Å, and diffracted to 1.92 Å. Here, we solved the YisP structure by using two Hg-containing derivatives (BsYisP has only 23% sequence identity with *S. aureus* CrtM and the structure could not be solved by molecular replacement). Full data collection and refinement statistics are given in Table 4.1. We obtained one structure (Figures 4.1b,c) with a molecule bound in the active site region, most likely a polyethylene glycol (PEG) from the crystallization buffer and not a bound substrate or product isoprenoid since as can be seen in the electron density (Figure 4.2), there are no methyl substituents and no planar groups. Numerous attempts to obtain structures of YisP with FPP, S-*thiolo*-FPP, S-*thiolo*-GGPP or farnesol ligands were unsuccessful.

The YisP molecule does indeed adopt the ε-fold expected, Figure 4.1b, but the DDLDE motif is located some 20 Å outside of the active site region found in the CrtM and SQS structures (Figures 4.1d, e) while the AASAD motif superimposes on the first Asp-rich motif found in these proteins, as does the second Asp-rich motif. A possible implication of these results is that YisP might not function as a head-to-head prenyl synthase since the first Asp residue found in CrtM and SQS is actually an alanine in YisP, and the DDLDE motif is well outside the active site region found in CrtM and SQS. There is one main pocket in the YisP structure with a total volume (calculated using PocketFinder[16]) of ~1000 $Å^3$. This pocket (Figure 4.1b, black mesh) is occupied by the single PEG molecule, Figure 4.1c, which is located in the S2 (prenyl donor) site

identified previously in CrtM and SQS[17]. Notably, this pocket volume is much smaller than that found in CrtM protein structures (~1600 $Å^3$, on average, Figure 4.1d) or HsSQS (~1900 $Å^3$, on average, Figure 4.1e), suggesting the possibility that the YisP product might be smaller than that the C_{30} species produced by CrtM and SQS. We next sought evidence for formation of head-to-head (squalene or squalene-like) condensation products since only phosphate release was detected in earlier work[10]. We carried out reactions with many combinations of isopentenyl diphosphate (IPP) /dimethylallyl diphosphate (DMAPP)/geranyl diphosphate (GPP)/FPP/geranylgeranyl diphosphate (GGPP) with Mg^{2+}, in the presence or absence of NADH/NADPH, but found no evidence for condensation products (by radio-TLC or by GC-MS) under conditions where large amounts of squalene/dehydrosqualene were readily detected with SQS or CrtM. We also used several different detergents (Triton X-100, Tween 80 and IGEPAL), but were again unable to detect any squalene/dehydrosqualene (or related) products. Nevertheless, consistent with earlier work, we did find that there was phosphate release using the preferred substrate, FPP, as a substrate, but the reaction rate was only minimally affected by NADH or NADPH, Figure 4.3a. The product found was, however, the sesquiterpene alcohol farnesol, as determined by radio-TLC using [^{14}C]-FPP (produced from GPP and [^{14}C]-IPP, Figure 4.3b). Plus, FPP reacted 2.5 times faster than did GPP and 6 times faster than did DMAPP, and GGPP is not active, in accord with results reported previously.[10] These results thus indicate that YisP acts as a phosphatase, not as a terpene synthase.

Site-directed mutagenesis results indicate that mutation of A54 and A55 in the AASAD motif to D had only a small effect on phosphate-release activity (Figure 4.4) and again no condensation products were formed, but D58A, D167A and D171A mutants had only background levels of activity, Figure 4.4, indicating a role in catalysis. These results are consistent with the x-ray crystallographic result that the DDLDE motif is not in the active site region seen in CrtM or SQS, and the 1st Asp in the DXXXD first Asp-rich motif found in CrtM and SQS in YisP is an Ala. Mutations at this position in SQS are known to abrogate activity and in SQS, this Asp is very highly conserved.[18,19] Therefore, YisP adopts the ε-fold seen in SQS/CrtM but acts as a phosphatase, raising the question

as to the origin of the pigmentation reported earlier[10] and the function of the possible YisP biological product, farnesol.

B. subtilis was reported in early literature to not produce carotenoids[20], consistent with our observation that YisP-the only "head-to-head" prenyl synthase-like enzyme in the *B. subtilis* genome, is a phosphatase. The question is, then, the nature of the pigmentation seen in *B. subtilis* under some growth conditions (light exposure). As shown in Figure 4.3c, the UV-VIS spectrum of a *B. subtilis* cell culture supernatant has two main features: a broad absorption centered at ~490 nm, and a three-peak absorption at shorter wavelengths (343, 364, 385 nm). The red pigment (λ_{max} ~ 490 nm) has previously been shown to arise in *B. subtilis* from the complex formed between Fe^{3+} and pulcherriminic acid[21-23], Scheme 4.1. The triple-peak absorption is suggestive of a carotenoid spectrum, but carotenoids absorb at much longer wavelengths, and the actual absorption wavelengths suggest a hexaene fragment, $(CH=CH)_6$, and is known in *B. subtilis* to arise from the eponymous bacillaene and bacillaene B3[24-26]. There is, therefore, no evidence for squalene/polyene-derived carotenoid pigments in *B. subtilis* and earlier conclusions to the contrary are hereby revised. We then investigated possible effects of the YisP product, farnesol, on *B. subtilis* biofilm formation. We used the Δ*yisP* mutant which is unable to form biofilms and whose construction was reported earlier[10] and as shown in Figure 4.3d, upon addition of farnesol, biofilm formation is clearly visible (at right) in the Δ*yisP* mutant. Similar results were obtained in Msgg culture, Figure 4.5, and there were no effects on wild-type biofilm formation.

The question then arises as to the likely role of farnesol in biofilm formation, and why zaragozic acid blocks biofilm formation- both in *B. subtilis* as well as in *S. aureus*. In previous work, an important role for lipid micro-domains in biofilm formation in *B. subtilis* (as well as *S. aureus*) was demonstrated[10] so we hypothesized that farnesol might be involved since it is well known to affect lipid bilayer structure, as shown by both differential scanning calorimetry (DSC) and by solid-state ^{2}H-NMR spectroscopy.[27-30] Specifically, farnesol shifts and broadens phospholipid gel-to-liquid crystal phase transitions and decreases ΔH – as does cholesterol[31,32], involved in lipid micro-domain (or raft) structures in eukaryotes, as illustrated in Figure 4.6a with dipalmitoylphosphatidyl choline (DPPC). Interestingly, we find that staphyloxanthin also

shifts and broadens the gel-to-liquid crystal phase transition and decreases ΔH. The polyene in staphyloxanthin cannot be readily "bent", due to the presence of 9 conjugated double bonds, in which case the molecule will act as a rigid body in the lipid bilayer (as does cholesterol) and its long/short side-chain structure suggests it may form an inter-digitated structure, Figure 4.6b. This rigidifying effect of staphyloxanthin has been proposed (based on diphenylhexatriene fluorescence depolarization results) to "subvert non-oxidative host defenses mediated by cationic peptides"[11] and could also block access of reactive oxygen species into *S. aureus*-an anti-virulence effect, and changes in lipid micro-domain structure also affect biofilm formation.[10] So, both the YisP product farnesol, as well as staphyloxanthin, affect lipid bilayer structure – effects that would be inhibited by zaragozic acid, as reported earlier.[10]

Staphyloxanthin is an isoprenoid virulence factor. In *M. tuberculosis*, tuberculosinol and the *iso*-tuberculosinols are also isoprenoid virulence factors, blocking phagolysosome maturation[33] as well as the uptake of opsonized zymosan particles into macrophages.[34] Essentially 100% of the tuberculosinol/*iso*-tuberculosinols in *M. tuberculosis* reside within the cell membrane[35], which strongly suggests that their role as virulence factors could also be to modulate lipid membrane structure. The chemical structures of the tuberculosinols suggest that they might have cholesterol-like effects on bilayer domain structure and indeed, as shown in Figure 4.6b, we do see that tuberculosinol again shifts and broadens the gel-to-liquid crystal phase transition (of DPPC) in a similar manner to that seen with farnesol, cholesterol and staphyloxanthin.

4.4 Conclusions

We report the first x-ray structure of the YisP protein, from *Bacillus subtilis*. The fold is very similar to that seen in *S. aureus* CrtM or human SQS. There are two "DDXXD"-like motifs, but one is ~20Å from the active site seen in SQS and CrtM. We find that YisP is a phosphatase, producing farnesol (from farnesyl diphosphate). The pigments and polyene formed by *B. subtilis* are not poly-isoprenoids, rather, the ferric iron complex of pulcherriminic acid, and bacillaenes. We also show that farnesol, the *S. aureus* virulence factor staphyloxanthin, as well as the *M. tuberculosis* tuberculosinol virulence factor, all affect lipid bilayer structure, shifting and broadening the main gel-to-liquid crystal phase

transition temperature, a cholesterol-like effect that we propose could play a role in their activity, in cells: modulating micro-domain (raft) structure.

4.5 Materials and Methods

Crystallization and Data Collection for BsYisP

BsYisP was cloned, expressed and purified as described previously.[15] Wild-type BsYisP was first crystallized by using the Grid Screen PEG/LiCl screen kit (Hampton Research, Aliso Viejo, CA) and the sitting-drop vapor diffusion method. The reservoir solution (No. C6) contained 1 M LiCl, 0.1 M Bicine, pH 9.0 and 20% w/v polyethylene glycol 6000. Better crystals were obtained by optimizing the crystallization condition to 1 M LiCl, 0.1 M Bicine, pH 9.0 and 18-20% w/v polyethylene glycol 6000 at 25 ^{o}C for 2 days. To prepare heavy-atom derivatives, the Hg-containing reagents of Heavy Atom Screen Hg (Hampton Research) were used. Cryoprotectant solutions (1 M LiCl, 0.1 M Bicine, pH 9.0, 20% w/v polyethylene glycol 6000 and 10% glycerol) containing 2 mM Hg derivatives were used for soaking the wild-type crystals for about 1 hour.

The X-ray diffraction datasets were collected at beam lines BL13B1 of the National Synchrotron Radiation Research Center (NSRRC, Hsinhu, Taiwan) and processed by using the HKL2000 program[36]. Prior to structural refinement, 5 % randomly selected reflections were set aside for calculating R_{free} as a monitor. The MIR datasets of mercury-containing derivatives were collected at a wavelength of 0.97622 Å (BL13B1). Combination of datasets from different Hg-derivative crystals with the “native” dataset from the wild-type crystal using SOLVE and RESOLVE[37] improved the figure of merit (FOM) values from 0.24 to 0.41, the Z-scores from 8.56 to 18.51, and the number of auto-built amino acid residues up to 177. The best results were obtained using mersalyl acid and mercuric acetate derivatives. Statistics of the two Hg-derivative data sets and the MIR phasing are summarized in Table 4.1. The model and map were further improved by computational refinement using Refmac5[38]. Finally, a complete model (but missing residues 57-81) with most side-chains was built using ARP/wARP[39]. Structural refinements were carried out using CNS[40] and Coot[41]. Data collection and refinement statistics are summarized in Table 4.1. All diagrams of protein structures were prepared by using PyMol (http://www.pymol.org/).

Mutant Preparation

The mutants were prepared with a BsYisP–pET32 Xa/LIC vector as the template and forward primers 5'-GTATTGTCGTTTTGCCACACGGATGCCAGTGCAGACGAAAA GGTA-3' for A54D, 5'-TGCCACACGGCGGCCAGTGCAGCGGAAAAGGTATTGCC CGCATTT' for D58A, 5'-CAGCTGGTTCGCTTCTTAAGTGCGCTTGGCACAGATC AGCAAAAA-3' for D167A, 5'-TTCTTAAGTGATCTTGGCACAGCGCAGCAAAAA AACCGCATTCCC-3' for D171A, and 5'-TTGTCGTTTTGCCACACGGATGATAGTG CAGACGAAAAGGTATTG-3' for A54D/A55D, were used. The recombinant plasmids were transformed to *E. coli* BL21 trxB (DE3) and proteins were expressed and purified as for wild type BsYisP[15].

BsYisP/Human SQS Activity Assay and Product Determination

The reaction of FPP catalyzed by BsYisP was monitored at 360 nm by using a continuous spectrophotometric assay[42] for diphosphate release in 96-well plates with 200 μL reaction mixtures containing 400 μM 7-methyl-6-thioguanosine (MESG), 0.1 U/mL purine nucleoside phosphorylase (PNP) and inorganic diphosphatase, 100 μM FPP, 100 μM NADPH or NADH, 25 mM Tris·HCl (pH 7.5), 0.2% detergent, and 1 mM $MgCl_2$. Detergents tested were Triton X-100, Tween 80 and IGEPAL, with IGEPAL giving the best activity.

Product identification of BsYisP and human SQS was carried out using radio-TLC with ^{14}C-labeled FPP as substrate. ^{14}C-labeled FPP was synthesized by adding 100 μM GPP, 100 μM [^{14}C]IPP and 1 μM HsFPPS to 50 μL of 25 mM Tris·HCl (pH 7.5), 0.2% detergent, and 1 mM $MgCl_2$, and reacting at 37°C for 30 min. BsYisP or human SQS (and 100 μM NADPH) was then added, and the resulting reaction mixture incubated at 37°C for 2 hours. The reactions were then terminated by adding 500 μL NaCl saturated H_2O, and the product extracted with n-hexane. The n-hexane solution was then concentrated and spotted onto a C18 reverse phase TLC plate (Whatman Ltd, Maidstone, UK). The TLC plate was developed with acetone/methanol (2:8) and the product distribution analyzed by using a Storm 840 PhosphorImager and Quantity One (Bio-Rad, Hercules, CA).

Biofilm Formation Assay

Biofilm formation assays in *B. subtilis* were carried out as described previously[10].

Differential Scanning Calorimetry (DSC)

1,2-dihexadecanoyl-*sn*-glycerophosphoryl choline (DPPC) was obtained from Avanti Polar Lipids (Alabama, USA). Farnesol and cholesterol were obtained from Sigma-Aldrich (Missouri, USA). Tuberculosinol and staphyloxanthin were made as previously described.[6,43] 20 mol % farnesol, cholesterol, tuberculosinol or staphyloxanthin were co-dissolved with DPPC in chloroform. The organic solvent was removed under a stream of dry nitrogen and the remaining solvent traces removed *in vacuo*, overnight. 10 µL deionized water was added and the mixtures homogenized by hand. Weighed amounts of the mixtures were then sealed into solid sample inserts (stainless steel tubes). The DSC experiments were performed on a Microcal VP-DSC instrument. The scans covered the range from 10 to 60 °C at a scan rate of 90 °C/h. The DSC thermograms were analyzed using Origin 7.1 (Massachusetts, USA). Water vs. water scans were used for baseline correction.

4.6 Schemes, Charts, Tables and Figures

Scheme 4.1 Structures of compounds of interest. a) Biosynthesis of virulence factors, sterols and farnesol. b) A biofilm inhibitor and some *B. subtilis* pigments.

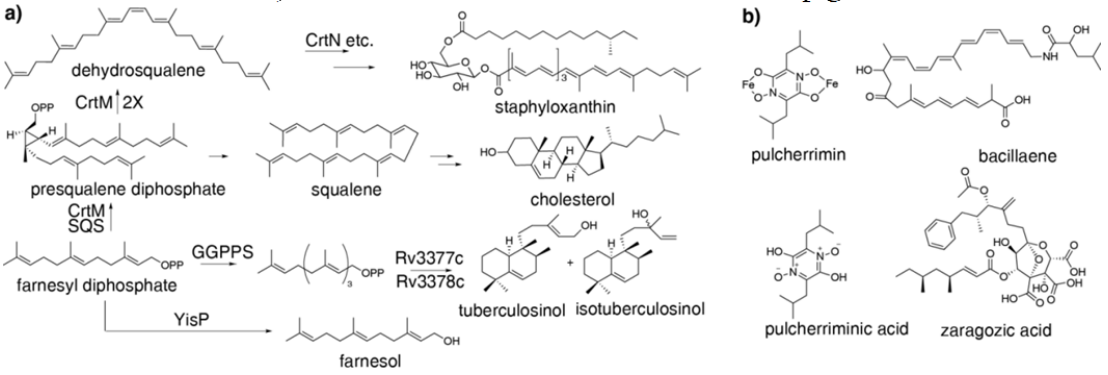


Table 4.1 Data collection and refinement statistics of BsYisP and Hg crystals. All positive reflections were used in the refinement. Values in parentheses are for the outermost resolution shells.

Name	BsYisP	$C_{13}H_{17}HgNO_6$	$Hg(OOCCH_3)_2$
Data collection			
Wavelength (Å)	0.97622	0.97622	0.97622
Resolution (Å)	25.00-1.92 (1.99-1.92)	25.00-2.34 (2.42-2.34)	25.00-2.92 (3.02-2.92)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell			
a (Å)	43.97	43.95	44.38
b (Å)	77.58	77.58	77.60
c (Å)	91.38	91.46	91.55
No. of measured reflections	166594 (16746)	84584 (8379)	45046 (4428)
No. of unique reflections	24603 (2427)	13675 (1330)	7249 (692)
Completeness (%)	99.9 (100.0)	99.9 (99.9)	100.0 (100.0)
R_{merge} (%) [a]	6.9 (48.2)	10.5 (49.6)	11.8 (47.6)
Mean $I/\sigma(I)$	32.7 (5.5)	19.6 (4.5)	16.8 (4.8)
Redundancy	6.8 (6.9)	6.2 (6.3)	6.2 (6.4)
Phasing			
No. of sites		3	
Z-score		18.51	
Figure of merit		0.74	
Refinement			
No. reflections used	23740 (2178)		
R_{work} (%)	21.1 (26.2)		
R_{free} (%)	23.3 (30.7)		
Geometry deviations			
Bond lengths (Å)	0.013		
Bond angles (^{o})	2.08		
No. of atoms / Mean B-values ($Å^2$)			
Protein atoms	2007 / 29.3		
Water molecules	219 / 39.0		
PEG molecules	9 / 48.1		
Ramachandran plot (%)			
Most favored	97.1		
Additionally allowed	2.5		
Disallowed	0.4		
PDB ID code	3WE9		

[a] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl)-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

Figure 4.1 YisP, CrtM and SQS sequences and structures. a) ClustalW alignment of BsYisP, SaCrtM and HsSQS showing 1st and 2nd Asp-rich domains and the DDLDE domain in BsYisP. b) X-ray structure of YisP (PDB ID code 3WE9) showing the conserved SARM, the DDLED domain, and the ~1000 Å^3 pocket. c) As b) but zoomed in to show the PEG ligand in S2, the SARM, and the location of the missing FARM. Structure superimposed on the S-thiolo-farnesyl diphosphates bound to CrtM (PDB ID code 2ZCP) showing PEG in the S2 site. d) CrtM liganded structure (PDB ID code 3ADZ) showing (canonical) ~ 1600 Å^3 pocket for S1+S2. The FARM and SARM are both conserved in the head-to-head prenyl synthases. e) Human SQS structure (PDB ID code 3LEE) showing larger pocket volume (~1900 Å^3 on average) due to the need for NADPH to bind in SQS but not in CrtM. The smaller pocket in YisP suggests a smaller product than that produced by CrtM or SQS.

a)

```
BsYisP   LPLKQRQASWAVLSFCH-TAASADEK-VLPAFEAKADHVYQR-----------TNNGKQHLWKAFDHAYRTFTLESEPFREFIAAQ--KE 112
SaCrtM   LPEDQRKAVWAIYAVCRKIDDSIDVY-GDIQFLNQIKEDIQSIEKYPYE--HHHFQSDRRIMMALQHVAQHKNIAFQSFYNLIDTVYKDQ 115
HsSQS    LDGEMRNAVCIFYLVLRALDTLEDDMTISVEKKVPLLHNFHSFLYQPDWRFMESKEKDRQVLEDFPTISLEFRNLAEKYQTVIADICRRM 150
         *  . * *   .  . :     *            .  :             : .  :   :          :  .*

BsYisP   DAKPYDDLDELLM-------YAYRTGGAAGLMLLPILTR-----RKQDQLKQAAVSLGLAIQLVRFLSDLGTDQQKNR--IPRQVMQQFG 186
SaCrtM   HFTMFETDAELFG-------YCYGVAGTVGEVLTPILSD-----HETHQTYDVARRLGESLQLINILRDVGEDFDNERIYFSKQRLKQYE 193
HsSQS    GIGMAEFLDKHVTSEQEWDKYCHYVAGLVGIGLSRLFSASEFEDPLVGEDTERANSMGLFLQKTNIIRDYLEDQQGGREFWPQEVWSRYV 240
         : .       *.   .*  *    ::                 *  :*  :*  .:: *   *   *   .::  .:
```

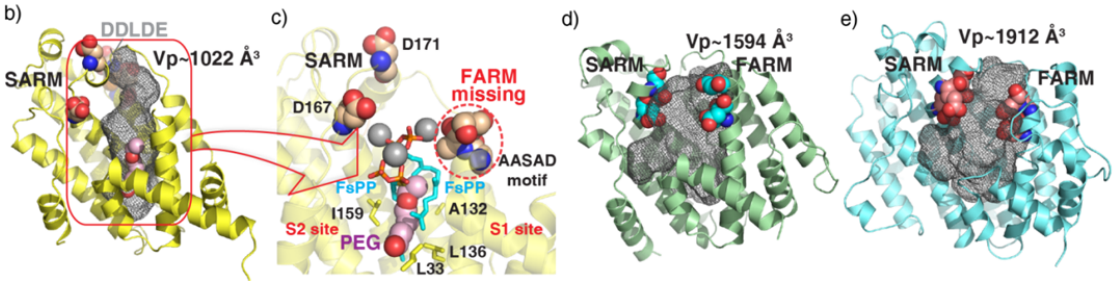


Figure 4.2 Electron density (contoured at 2σ (red) and 1σ (grey)) for the ligand in the YisP S2 site, modeled as a PEG fragment.

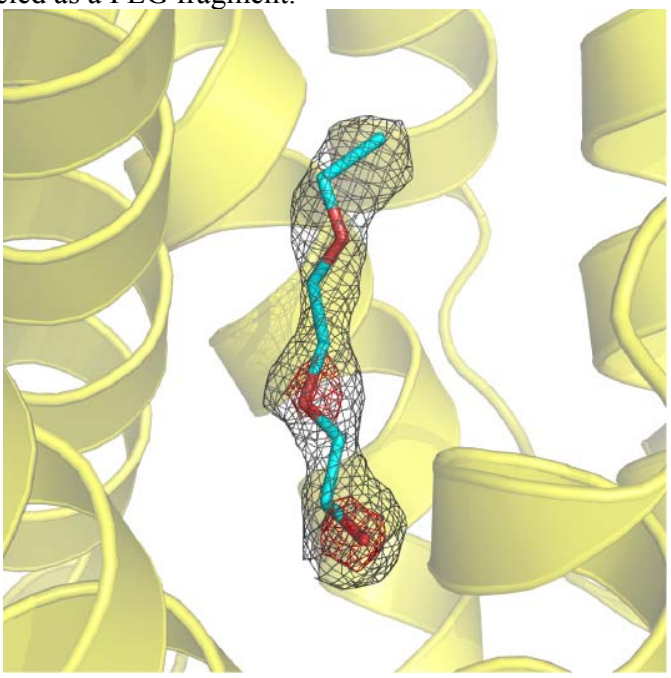


Figure 4.3 YisP activity, pigment formation and biofilm formation in B. subtilis. a) NADPH or NADH shows little effect on K_{cat} of YisP. b) FPP forms farnesol with YisP but squalene with SQS+NADPH. The top trace is the control (^{14}C-FPP, no YisP). c) UV-VIS of *B. subtilis*. Cells were grown in the dark and produced a dark pink pigment (pulcherrimin); the 3-band peak is from the bacillaenes. d) Biofilm formation requires the production of an extracellular matrix, which is correlated with the amount of wrinkles observed in the colony[10]. Treatment with 500 μM farnesol restores biofilm formation in the Δ*yisP* mutant.

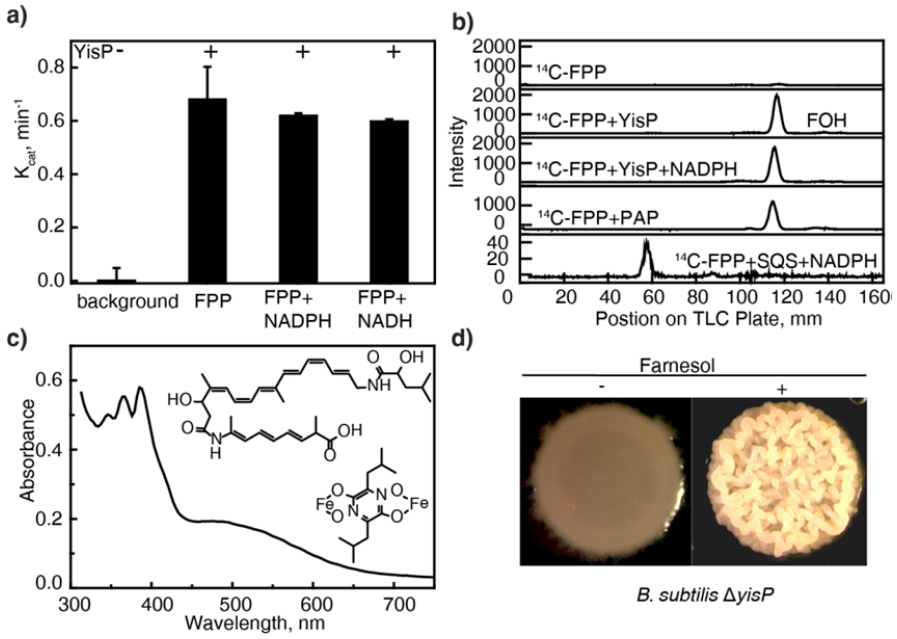


Figure 4.4 Relative actives of wild-type and mutant YisP.

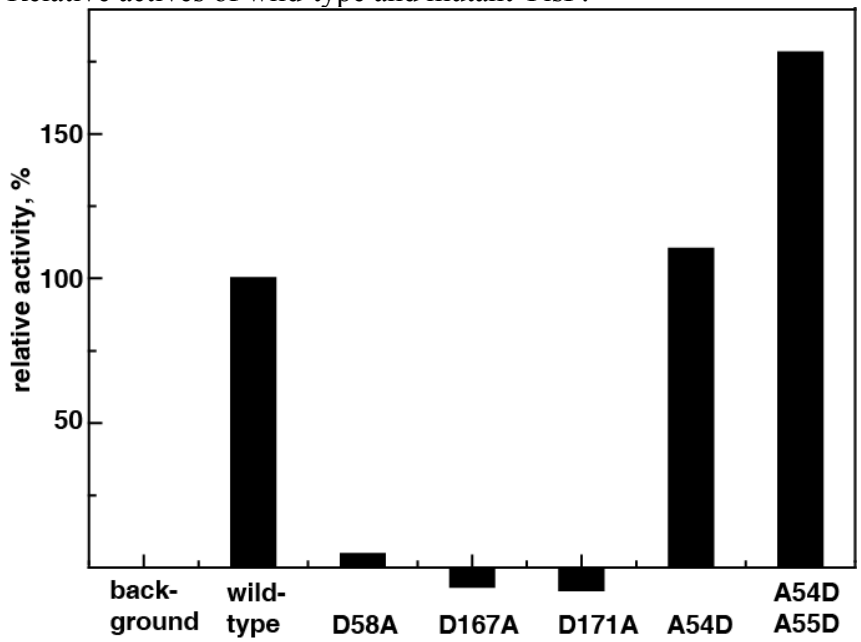


Figure 4.5 Effects of treatment with 500 μM farnesol on biofilm formation in wild-type *B. subtilis* and Δ*yisP* mutant in Msgg culture.

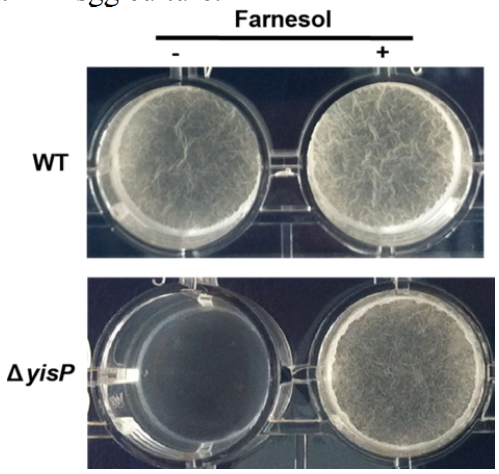
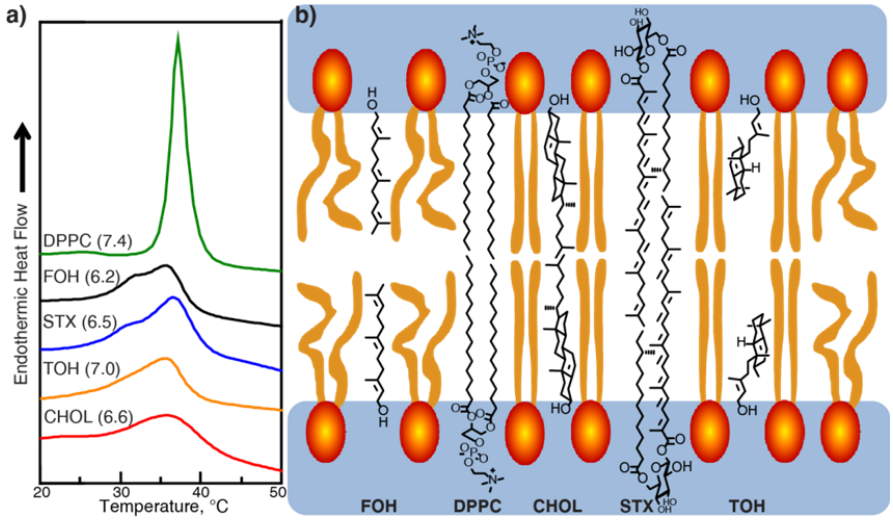


Figure 4.6 Effects of virulence factors/raft modulators on lipid bilayers. a) DSC thermograms for 1,2-dipalmitoyl-sn-glycerophosphatidyl choline (DPPC, structure shown in b)) in the presence of 20 mole% additive. The additives were farnesol (FOH); staphyloxanthin (STX); tuberculosinol (TOH) and cholesterol (CHOL). The numbers in parentheses are the ΔH for the main gel-to-liquid crystal phase transitions (in kcal/mole). Samples were in excess water. b) Cartoon showing proposed organization of virulence factors/raft modulators in membranes. The black stick DPPC structure serves as a "ruler" for CHOL, FOH, STX and TOH. The orange cartoons represent ordered or disordered DPPC molecules.



4.7 References

(1) Oldfield, E.; Lin, F. Y. *Angew. Chem. Int. Ed.* **2012**, 51, 1124.
(2) Oldfield, E. *Acc. Chem. Res.***2010**, 43, 1216.
(3) Kourounakis, A. P.; Katselou, M. G.; Matralis, A. N.; Ladopoulou, E. M.; Bavavea, E. *Curr. Med. Chem.* **2011**, 18, 4418.
(4) Urbina, J. A. *Mem. Inst. Oswaldo Cruz* **2009**, 104, 311.
(5) Wieland, B.; Feil, C.; Gloriamaercker, E.; Thumm, G.; Lechner, M.; Bravo, J. M.; Poralla, K.; Gotz, F. *J. Bacteriol.* **1994**, 176, 7719.
(6) Pelz, A.; Wieland, K. P.; Putzbach, K.; Hentschel, P.; Albert, K.; Gotz, F. *J. Biol. Chem.* **2005**, 280, 32493.
(7) Liu, G. Y.; Essex, A.; Buchanan, J. T.; Datta, V.; Hoffman, H. M.; Bastian, J. F.; Fierer, J.; Nizet, V. *J. Exp. Med.* **2005**, 202, 209.
(8) Clauditz, A.; Resch, A.; Wieland, K. P.; Peschel, A.; Gotz, F. *Infect. Immun.* **2006**, 74, 4950.
(9) Resch, A.; Rosenstein, R.; Nerz, C.; Gotz, F. *Appl. Environ. Microbiol.* **2005**, 71, 2663.
(10) Lopez, D.; Kolter, R. *Genes. Dev.* **2010**, 24, 1893.
(11) Mishra, N. N.; Liu, G. Y.; Yeaman, M. R.; Nast, C. C.; Proctor, R. A.; McKinnell, J.; Bayer, A. S. *Antimicrob. Agents. Chemother.* **2011**, 55, 526.
(12) Liu, C.-I.; Jeng, W.-Y.; Chang, W.-J.; Ko, T.-P.; Wang, A. H. J. *J. Biol. Chem.* **2012**, 287, 18750.
(13) Kelley, L. A.; Sternberg, M. J. E. *Nat. Protoc.* **2009**, 4, 363.
(14) Christianson, D. W. *Chem.Rev.* **2006**, 106, 3412.
(15) Hu, Y.; Jia, S.; Ren, F.; Huang, C. H.; Ko, T. P.; Mitchell, D. A.; Guo, R. T.; Zheng, Y. *Acta Crystallogr. F* **2013**, 69, 77.
(16) Hendlich, M.; Rippmann, F.; Barnickel, G. *J. Mol. Graph. Model.* **1997**, 15, 359.
(17) Liu, C. I.; Liu, G. Y.; Song, Y.; Yin, F.; Hensler, M. E.; Jeng, W. Y.; Nizet, V.; Wang, A. H.; Oldfield, E. *Science* **2008**, 319, 1391.
(18) Gu, P. D.; Ishii, Y.; Spencer, T. A.; Shechter, I. *J. Biol. Chem.* **1998**, 273, 12515.
(19) Lin, F. Y.; Liu, C. I.; Liu, Y. L.; Zhang, Y.; Wang, K.; Jeng, W. Y.; Ko, T. P.; Cao, R.; Wang, A. H.; Oldfield, E. *Proc. Natl. Acad. Sci. USA* **2010**, 107, 21337.
(20) Clejan, S.; Krulwich, T. A.; Mondrus, K. R.; Setoyoung, D. *J. Bacteriol.* **1986**, 168, 334.
(21) Canaleparola, E. *Archiv Fur Mikrobiologie* **1963**, 46, 414.
(22) Uffen, R. L.; Canale-Parola, E. *J. Bacteriol.* **1972**, 111, 86.
(23) Kupfer, D. G.; Uffen, R. L.; Canalepa.E *Archiv Fur Mikrobiologie* **1967**, 56, 9.
(24) Kudryashova, E. B.; Vinokurova, N. G.; Ariskina, E. V. *Appl. Biochem. Microbiol.* **2005**, 41, 486.
(25) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straightt, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *Proc. Natl. Acad. Sci. USA* **2007**, 104, 1506.
(26) Moldenhauer, J.; Goetz, D. C. G.; Albert, C. R.; Bischof, S. K.; Schneider, K.; Suessmuth, R. D.; Engeser, M.; Gross, H.; Bringmann, G.; Piel, J. *Angew. Chem. Int. Ed.* **2010**, 49, 1465.
(27) Bondar, O. P.; Melnykovych, G.; Rowe, E. S. *Chem. Phys. Lipids* **1994**, 74, 93.
(28) Rowat, A. C.; Davis, J. H. *Biochim. Biophys. Acta* **2004**, 1661, 178.
(29) Funari, S. S.; Prades, J.; Escriba, P. V.; Barcelo, F. *Mol. Membr. Biol.* **2005**, 22, 303.

(30) Rowat, A. C.; Keller, D.; Ipsen, J. H. Biochimica Et *Biochim. Biophys. Acta* **2005**, 1713, 29.
(31) Ladbrook.Bd; Williams, R. M.; Chapman, D. *Biochim. Biophys. Acta* **1968**, 150, 333.
(32) Oldfield, E.; Chapman, D. *FEBS Lett.* **1972**, 23, 285.
(33) Pethe, K.; Swenson, D. L.; Alonso, S.; Anderson, J.; Wang, C.; Russell, D. G. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 13642.
(34) Hoshino, T.; Nakano, C.; Ootsuka, T.; Shinohara, Y.; Hara, T. *Org. Biomol. Chem.* **2011**, 9, 2156.
(35) Prach, L.; Kirby, J.; Keasling, J. D.; Alber, T. *FEBS Lett.* **2010**, 277, 3588.
(36) Otwinowski, Z.; Minor, W. *Method Enzymol.* **1997**, *276*, 307.
(37) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D* **2010**, *66*, 213.
(38) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D* **1997**, *53*, 240.
(39) Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A. *Nat. Protoc.s* **2008**, *3*, 1171.
(40) Emsley, P.; Cowtan, K. *Acta Crystallogr. D* **2004**, *60*, 2126.
(41) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D* **1998**, *54*, 905.
(42) Webb, M. R. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4884.
(43) Nakano, C.; Ootsuka, T.; Takayama, K.; Mitsui, T.; Sato, T.; Hoshino, T. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 75.

Chapter 5: Structure, function and inhibition of *ent*-kaurene synthase from *Bradyrhizobium japonicum*

5.1 Notes and Acknowledgements

W. Liu, Y. Zheng, C.-H. Huang and X. Feng cloned, expressed and purified BjKS proteins. W. Liu, Y. Zheng, C.-C. Chen, T.-P. Ko and R.-T. Guo carried out crystallization, data collection, structure determination and refinement. X. Feng and S. Bogue characterized the enzymatic activity of BjKS and carried out mutagenesis study. C. Nakano and T. Hoshino provided *ent*-copalyl diphosphate. Y. Cui, J. Li, X. Feng, E. Oldfield and R.-T. Guo analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by grants from the National Basic Research Program of China (2011CB710800 and 2011CBA00805), the National Natural Science Foundation of China (31200053 and 31300615), and by the United States Public Health Service (NIH grants GM065307 and CA158191). X.F. was supported by a Pre-doctoral Fellowship from the American Heart Association, Midwest Affiliate (13PRE14510056). We thank the National Synchrotron Radiation Research Center of Taiwan for beam-time allocation and data collection assistance.

This chapter was reproduced in part with permission from W. Liu, X. Feng, Y. Zheng, C.-H. Huang, C. Nakano, T. Hoshino, S. Bogue, T.-P. Ko, C.-C Chen, Y.Cui, J, Li, E. Oldfield, R.-T. Guo, paper submitted to *Chemical Science* on March 10, 2014. W. Liu, X. Feng and Y. Zheng contributed equally to this work.

5.2 Introduction

Terpenes (or terpenoids) represent the largest class of small molecules on the planet[1] and most are synthesized by enzymes that contain one or more of six main types of fold: α, β, γ, δ, ε and ζ[2]. About 20% of terpenes are diterpenes, molecules that contain a core of 20 carbon atoms, and most of these are made by plants and fungi. Diterpenes have, however, also been found in bacteria[3-11], and these compounds are of interest since some have activity as anti-infective and anti-cancer drug leads, virulence factors, as well as plant growth hormones. There are two different classes of terpene cyclases, Class I and

Class II, defined by the amino acid motifs that are essential for catalysis[12-14]. Class I cyclases contain two conserved motifs, **D**DXX**D**, and (**N,D**)DXX(**S,T,G**)XXX(**E,D)** (hereafter referred to as NTE), in which the boldface residues generally bind to the three Mg^{2+} that facilitate ionization of an isoprenoid diphosphate group, generating a reactive carbocation intermediate. In contrast, Class II cyclases typically contain only a DX**D**D motif in which the boldface aspartic acid protonates an isoprenoid double bond or an oxirane to, once again, generate a reactive carbocation intermediate. In some cases, both Class I as well as Class II motifs are present and these molecules are bi-functional or in one reported case, tri-functional[15]. For example, abietadiene synthase from *Abies grandis* (AgABS) is bi-functional and converts geranylgeranyl diphosphate (GGPP) to copalyl diphosphate (CPP) in a Class II reaction and then CPP is converted to abietadiene in a Class I reaction, Scheme 5.1a. Similarly, the moss *Physcomitrella patens* utilizes a trifunctional CPPS/KS (PpCPPS/KS) to convert GGPP to *ent*-CPP in a Class II reaction and then to *ent*-kaur-16-ene (20%) and 16α-hydroxy-*ent*-kaurane (80%) via a Class I reaction with, in the case of kauranol, carbocation quenching[16], Scheme 5.1b.

Here, we report the first structure of a bacterial diterpene cyclase, *ent*-kaurene synthase from the soil bacterium *Bradyrhizobium japonicum* (BjKS; gene # blr2150)[4]. BjKS catalyzes the cyclization of *ent*-CPP which is produced by *ent*-copalyl diphosphate synthase (*ent*-CPPS; blr2149) from GGPP (Scheme 5.1c), and one possible mechanism of action [17] for the *ent*-kaurene synthase reactions is shown in Scheme 5.1c. The *ent*-KS reactions proceed via initial diphosphate loss from *ent*-CPP (catalyzed by Mg^{2+}) and cyclization to form the 8-carbonium ion, the pimaren-8-yl cation, which then undergoes secondary cyclization with Wagner-Meerwein migration of C-12 to C-16 to form the *ent*-kauran-16-yl cation. This cation can then loose a proton to form *ent*-kaur-16-ene, or in *P. patens* the *ent*-kauranyl cation can also be quenched by a water molecule to form 16-α-hydroxy-*ent*-kaurane, Scheme 5.1b[16].

Structurally, Class I proteins contain a catalytic α domain while Class II proteins contain two domains, βγ, and with bi-functional species such as abietadiene synthase, all three domains (αβγ) are present[18]. In previous work[19] it was proposed that these αβγ domain proteins might have arisen by the fusion of the genes of α+βγ proteins in ancestral, perhaps soil dwelling, bacteria, so determining the structures of bacterial α

(and βγ) cyclases is of interest not only from a mechanism of action perspective, but also from an evolutionary one. Is, for example, the structure of BjKS (a predicted[19] α domain protein) most similar to that of the α domains seen in modern plant terpene cyclases? To investigate this and other questions we determined the structure of BjKS in the presence of its *ent*-CPP substrate, as well as apo- and inhibitor-bound structures, and we used site-directed mutagenesis to probe its catalytic activity. In addition, we compared the BjKS structure with the structures of diverse other terpene cyclases, and synthases.

5.3 Results and Discussions

Structure of apo-*ent*-kaurene synthase

We cloned, expressed and purified *ent*-kaurene synthase from *B. japonicum* then crystallized it and solved its structure using a SeMet-labeled protein (Supplementary Table 5.1). Full data acquisition and crystallographic details for two native (i.e. apo) crystals are given in Table 5.2. The final models were refined at a resolution of 1.9 – 2.0 Å using the data obtained with two apo crystals. A stereo-view of the apo-1 structure (PDB ID code 4KT9) is shown in Figure 5.1a from which it can be seen that the structure is mainly composed of twelve α helices, Figure 5.1a, as found in many other terpene cyclases as well as synthases such as farnesyl diphosphate synthase. There is, however, a missing loop (residues ~211-221) in both chains C and D in the apo-1 structure (PDB ID 4KT9) and this loop is also missing (that is, disordered) in all other structures (Table 5.3).

BjKS crystallizes as a dimer (Figure 5.1b) and there is a cluster of side-chains of hydrophobic residues (M235, Y232 and F240, cyan spheres) from both chains at the "bottom" of the dimer interface that form a hydrophobic core, Figure 5.1b, with a buried surface area of ~1000 $Å^3$. Further up (green spheres), the E244 side-chain interacts with the N-terminus of the helix starting at Y155 in the other subunit in the dimer, and salt bridges exist between the E157(A) and R250(B) side-chains, and between E157(B) and R250(A) (purple spheres). Using gel filtration, we find (Figure 5.1c and Figure 5.2) that BjKS also exists as a dimer in solution (M_r = 68.4 kDa; 61.9 kDa calc. from the sequence). However, as can be seen in Figure 5.1b the predicted catalytic motifs (DDXXD and NTE) are ~20Å from the dimer interface and unlike some other prenyl

synthases[20], there is no chain inter-digitation. This suggests that the presence of a dimer is not essential for catalysis.

The BjKS structure is generally similar to that found in the α domains (α_c[19]) of terpene cyclases from bacteria, fungi as well as plants: some representative examples are shown in Figure 5.3. Figure 5.3a shows a superposition between BjKS (red; PDB ID 4KT9) and pentalenene synthase (green; PDB ID 1HM7)[21] from the bacterium *Streptomyces* UC5319 where there is a 2.7 Å Cα rmsd (root mean square deviation) between Cα atoms over 204 residues; Figure 5.3b shows a superposition with 1, 8-cineole synthase (PDB ID 2J5C) from the flowering plant *Salvia fruticosa* (Sf-CinS1, blue), an "αβ" domain protein[19] in which the Cα rmsd in 2.4 Å over 195 residues, while Figure 5.3c shows a superposition with abietadiene synthase from the gymnosperm *Abies grandis* (AgABS; PDB ID 3S9V), an αβγ domain protein[18]. As can be seen in Figure 5.3c, the BjKS structure (red) is similar to the α domain of AgABS (yellow) and has a 2.6 Å Cα rmsd over 224 residues. The Cα rmsd value for the three other αβγ domain proteins: taxadiene synthase (PDB ID 3P5R); *ent*-copalyl diphosphate synthase (PDB ID 3PYB) and bisabolene synthase (PDB ID 3SDV) are 2.6 Å/201 residues, 3.5 Å/169 residues, and 2.9 Å/226 residues for taxadiene synthase, *ent*-copalyl diphosphate synthase and (the sesquiterpene) bisabolene synthase, respectively, and the close similarity to abietadiene synthase may be due to the fact that AgABS produces a C_{20} species that is structurally more similar to *ent*-kaurene than are the products of the three other αβγ proteins. Remarkably, however, there is only a 17% sequence identity between BjKS and AgABS.

We next sought to see how the BjKS structure compared with a broad range of α, β, γ, δ, ε, and ζ-fold proteins using their Q-scores [22]. The Q-score attempts to minimize difficulties than may arise by use of e.g. just Cα rmsd values because the numbers of aligned residues as well as the numbers of residues in target and query proteins, and their structural deviations, are all important. The Q-score is defined as [22,23]:

$$Q = \frac{N_{algn} * N_{algn}}{(1 + (rmsd/R_0)^2) * N_{res1} * N_{res2}}$$

where N_{algn} is the number of aligned residues, determined here by using the secondary structure matching (SSM) program[22]; N_{res1} and N_{res2} are the total numbers of residues in

the two aligned proteins whose structures are to be compared, rmsd in the Cα root mean square deviation of the Cα residues in the two proteins, and R_o is an empirical parameter, set (by the program) to 3 Å[22,23]. The Q-score is 1 if two structures are identical but obviously decreases as the rmsd increases, N_{algn} decreases, or if N_{res1} or N_{res2} are much larger than N_{algn}. As expected, Figure 5.4, the use of complete structures results in a clustering of BjKS with other bacterial and fungal cyclases (which are all small proteins, as is BjKS), while the plant αβ or αβγ proteins are very distant because of the large N_{res} in these systems. Q-scores for Figure 5.4 are shown in Table 5.4. However, since we are primarily interested in structural comparisons between α domains, it is logical to construct a second tree using just the highly α-helical domain-containing proteins to see where BjKS clusters: the α and αδ head-to-tail prenyl diphosphate synthases such as FPPS; the head-to-head synthases (which have an ε-fold) squalene and dehydrosqualene synthase (which utilize cyclic transition states/reactive intermediates), and the bacterial, fungal and plant cyclases (that contain α, αβ or αβγ domains). As shown in Figure 5.3d, using this subset of structures we see that BjKS now clusters with the plant α domains, while the other bacterial and fungal cyclases are separate, as are other α proteins such as farnesyl diphosphate synthase and geranylgeranyl diphosphate synthase. Q-scores for Figure 5.3d are shown in Table 5.5. Interestingly, using the Q-score metric the αβγ domain protein *ent*-CPPS from the plant *Arabidopsis thaliana* is rather distant from the other αβγ domain plant proteins (based on just the α domain structure) due, we propose, to the fact that the α domain in *ent*-CPPS is not catalytic in this protein, so its structure is less likely to be conserved.

There are, therefore, several pieces of information that are consistent with (or in support of) the α+βγ gene fusion hypothesis[19] for the origin of modern plant terpene cyclases. The BjKS structure most closely resembles the α domain structures in plant terpene cyclases, and as noted previously[4,19], in *B. japonicum*, the *ent*-CPPS and *ent*-KS genes overlap by a single nucleotide[4], a potential route to bifunctional species. Moreover, in primitive plants such as mosses and liverworts[16] there are bi-functional *ent*-KS enzymes, again suggesting a gene fusion origin for more modern plant terpene cyclases.

Mutagenesis and A Substrate-bound Structure

We next sought to investigate the catalytic mechanism of BjKS. We first used the BjKS and AgABS sequences to generate a set of 354 homologous proteins (using the JPRED3 program[24]) which was then used as input to the SCORECONS server[25] which ranks residues in terms of their essential nature: 1.000 is essential, 0 is non-essential. The top 10 most essential residues are shown in Figure 5.5a. We show in Figure 5.5b the locations of the top 10 non-structural (i.e. not Gly or Pro) residues that are most conserved, from which it can be seen that D75, D76, D79, R204 and D208 (red spheres) are close to the catalytic site seen in other terpene cyclases while the other residues are more distant. We elected to mutate D75, D79 and R204 (to D75A, D75C, D79C and R204A) to test whether they were essential for catalysis since D75 and D79 are part of the highly conserved (Figure 5.6) **D**DXXD motif found in many Class I cyclases, and R204 is also highly conserved, as part of the **RL**X(**N**,**D**)DXX(**S**,**TR**,**G**)XXX(**E**,**D**) (Figure 5.6) motif. As can be seen in Figure 5.5c and Table 5.6, mutation of D75, D79 or R204 (to Ala or Cys) greatly reduced activity to, on average, 1.7 ± 1.6%. Clearly, these residues are likely to be part of the active site, raising the question: where does the *ent*-CPP substrate bind?

In other articles[26] there have been reports of the binding of CPP isosteres (containing nitrogen) to terpene cyclases, but there have been no reports of *ent*-CPP (or CPP) bound to a terpene cyclase (or synthase). We therefore next determined the structure of the substrate *ent*-CPP (Scheme 5.1) bound to the BjKS D75C mutant. Full data acquisition and crystallographic details are given in Table 5.2. Electron densities are in Figure 5.7a, b. The BjKS structure with *ent*-CPP bound is shown in Figures 5.8a and 4b, and in Figure 5.8c, the BjKS-*ent* CPP structure is shown superimposed on that of 2-F-geranylgeranyl diphosphate (FGG) bound to the α domain of taxadiene synthase (TXS; PDB ID 3P5R). The 3 Mg^{2+} are from the taxadiene synthase structure. Four Asp residues (D75, D76, D79 and D208) in BjKS are located in very similar positions to the corresponding Asp residues in TXS (Cα rmsd of 0.8 Å for these 4 Asp residues), and are expected to coordinate to a $[Mg^{2+}]_3$ cluster which was, however, not observed in our BjKS crystals. Cα rmsd for these 4 Asp residues is 0.8 Å between BjKS and AgABS (3S9V), and 0.9 Å between BjKS and AgBIS (3SDQ). The hydrophobic side-chain of *ent*-CPP is buried in a

hydrophobic pocket composed of I36, L68, L71, F72, Y136, I166, A167, Y168 and N207, shown as the orange surface in Figure 5.8c. In this conformation, the *exo*-methylene group is poised (~4 Å distant) to interact with the carbocation formed on diphosphate ionization, forming the first initial reactive intermediate, the pimaren-8-yl ion (Scheme 5.1).

Inhibition by a Bisphosphonate

We also obtained the structure of an inhibitor-bound complex since such compounds might be of interest as leads for plant growth hormones targeting gibberellin biosynthesis. The structures of several bisphosphonates bound to the α domain in the αβγ protein bisabolene synthase have been reported[27], so we chose to investigate a bisphosphonate inhibitor, BPH-629 (Figure 5.9a), which has a $C_{19}O$ side-chain, comparable in size to the side-chain in *ent*-CPP. We found that BPH-629 inhibited BjKS with an IC_{50} = 9.5 μM, Figure 5.9a. We obtained the structure (at 2.34 Å resolution) shown in Figure 5.9b, and found the inhibitor occupies the same site as does *ent*-CPP, Figure 5.9c. Full data acquisition and crystallographic details are given in Table 5.2. Ligand electron densities are in Figure 5.7c. The bisphosphonate has hydrogen bonds from the bisphosphonate group to D75, S165, R204 and N207, Figure 5.9b, but just as with the *ent*-CPP complex, there is no Mg^{2+} observed. The lack of a $[Mg^{2+}]_3$ cluster in this case cannot be due to use of a D75C mutant (as used with the *ent*-CPP structure) since wild type protein was employed for the bisphosphonate structure. Reasons are not known.

Structural Basis for Formation of *ent*-kaurene and kauranol

The results described above delineate the active site in BjKS together with several residues (D75, D79 and R204) that are essential for catalysis. These residues comprise part of the conserved DDXXD/NTE motif involved in diphosphate loss to form carbocation transition states/reactive intermediates (Scheme 5.1c) that need to be protected from solvent attack in BjKS raising the question: which residues provide this protection? As noted above, in the moss *Physcomitrella patens*, there is in fact solvent attack in the PpCPPS/KS protein[16] which results in the formation of only 20% *ent*-kaurene but 80% 16-α-hydroxy-*ent*-kaurane and Kawaide *et al.*[15] identified a key residue

(A710, corresponding to L140 in BjKS) that enabled water access, quenching the kauranyl cation. A comparison between the two protein structures would be of interest since it would help clarify the BjKS mechanism of action, but the structure of the *P. patens* protein has not been reported. There is, however, a 45% sequence identity between PpCPPS/KS and the *A. grandis* abietadiene synthase and a Phyre2[28] structure prediction of the tri-functional *P. patens* protein is shown in Figure 5.10a. As expected, the protein is predicted to have an αβγ domain structure and the 45% conserved residues, shown in red in Figure 5.10a, are distributed over all three domains. When the *P. patens* active site structure prediction is superimposed on the BjKS structure (with *ent*-CPP from the BjKS structure) it is clear that BjKS contains larger residues (Y136 and L140) adjacent the *ent*-CPP ligand-binding site than those found in *P. patens* (Leu and Ala), and these larger residues (Figure 5.10b) prevent water from quenching the kauranyl cation (Scheme 5.1). The increase in side-chain volume is expected to be ~100 $Å^3$ (based on amino-acid volumes) and such steric differences might be particularly important in the presence of Mg^{2+}, because Class I terpenoid synthases typically undergo a structural transition from an open to a closed active-site conformation after the binding of three Mg^{2+} to the substrate's diphosphate group, helping to protect the reactive carbocation intermediates from premature quenching by bulk solvent.

In addition to Y136 and L140 protecting the carbocation intermediates, we propose that F72 plays a similar role since F72 is very close to the *ent*-CPP substrate, as shown in Figure 5.10c. The phenyl group in F72 could also be involved in cation-π stabilization of one or more reactive intermediates, a widely used catalytic strategy adopted by terpenoid cyclases[29-31]. Based on the x-ray structural results and structure comparison, it thus appears that Y136 and L140 can protect carbocations from solvent attack on one side of the intermediate(s), while F72 protects from solvent attack from the other face, as shown in Figure 5.10c.

5.4 Conclusions

The results presented above are of interest for several reasons. First, we report the first X-ray crystallographic structure of a bacterial diterpene cyclase. The structure is most similar to that found in the α domain in plant diterpene cyclases. Second, we report

the first structure of a diterpene diphosphate (*ent*-copalyl diphosphate) bound to a diterpene cyclase. The structure is of interest because it helps define the diterpene cyclase substrate-binding site using a native substrate, and the diphosphate group is found to be close to three residues that we find (using site-directed mutagenesis) are essential for catalysis. Third, we obtained the structure of an aromatic bisphosphonate inhibitor. The bisphosphonate binds in or very close to the *ent*-CPP diphosphate-binding site, while the dibenzofuran side-chain occupies basically the same site as does the decahydronaphthalene ring in *ent*-CPP: such compounds might find use as plant growth regulators. Fourth, the results of computational modeling of a *P. patens* tri-functional KS together with a comparison with the BjKS structure provide strong evidence in support of the roles of L140, Y136 and F72, in carbocation protection in the bacterial enzyme.

5.5 Materials and Methods

Cloning, expression and purification of *ent*-kaurene synthase from *Bradyrhizobium japonicum* (BjKS).

The gene encoding KS from *Bradyrhizobium japonicum* USDA 110 DNA was amplified by polymerase chain reaction (PCR) with a forward primer 5'-GGTATTGAGGGTCGCGAGAATCTTTATTTTCAGGGCGCTGGTGCTGGTGCTAT GAT-3' and a reverse primer 5'-AGAGGAGAGTTAGAGCCGTTATTAGGCCGGCGC GCGCTGGCCTCCCCTCAC-3', then cloned into the pET32 Xa/LIC vector. A tobacco etch virus protease (TEV protease) cutting sequence (ENLYFQG) was added before the KS gene. The recombinant plasmids were transformed to *E. coli* BL21*trxB* (DE3) and protein was induced with 0.5 mM isopropyl-thiogalactopyranoside (IPTG) at 16 °C for 20 hours. The cell lysate was harvested by centrifugation and the resulting supernatant loaded onto a Ni-NTA column equilibrated in buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl, and 20 mM imidazole. His-tagged KS was eluted using an imidazole gradient (20-250 mM). The protein solution was then dialyzed against a buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl and subjected to TEV protease digestion, to remove the thioredoxin and His tags. The mixture was then again passed through a Ni-NTA column and untagged KS eluted with a 5 mM imidazole-containing buffer. Protein was further purified by FPLC using a DEAE column, 25 mM Tris, pH 8.0, and a 0-500 mM NaCl

gradient. The protein eluted at ~170 mM NaCl and was then concentrated to 10 mg/mL in 25 mM Tris, pH 7.5, and 150 mM NaCl. Expression of selenomethionine-substituted (SeMet) KS was carried out as described previously[32] and the purification procedures were the same as those used for the wild-type enzyme, but the DEAE column was omitted. The protein was then concentrated to 6.4 mg/mL in 25 mM Tris, pH 7.5 and 150 mM NaCl.

Gel filtration chromatography

Gel filtration chromatography with a HiLoad 16/60 Superdex 200 column (column volume V_c= 121 mL) was used to determine the oligomeric state of wild type BjKS. The column was washed with 2 column volumes of 25 mM Tris, pH 7.5, and 150 mM NaCl buffer to equilibrate before sample injection, and 1 mL Blue Dextran 2000 (3 mg/mL) was injected to obtain the column void volume (V_o = 48 mL). A LMW calibration kit was used (GE Healthcare, 28-4038-41) which contains aprotinin (6.5 kDa), ribonuclease A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa) and Blue Dextran 2000 (>2000 kDa) was used. The above protein standards and BjKS were all dissolved in 25 mM Tris, pH 7.5, and 150 mM NaCl, at 1 mg/mL for ovalbumin and 0.8 mg/mL for the others. 1 mL of sample was loaded to the column at a flow rate of 0.5 mL/min. K_{av} is calculated as a function of elution volume V_e, column volume V_c and column void volume V_o as shown below.

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

A calibration curve was obtained by plotting K_{av} for the protein standards versus the logarithm of their molecular weights, and the molecular weight of BjKS was calculated to be 61.0 kDa using an elution volume V_e =74.7 mL and the standard curve.

Crystallization, data collection and structure determination of BjKS.

The wild-type KS protein containing 1 mM $MgCl_2$ and 10 mM DTT was first crystallized by using the Index Kit (Hampton Research, Aliso Viejo, CA) and the sitting-drop vapor diffusion method. The reservoir solution (No. 26) contained 1.1 M ammonium tartrate dibasic, pH 7.0. Better crystals were obtained by optimizing the

reservoir composition to 0.7 M ammonium tartrate dibasic, pH 7.0 and 2-4% w/v PEG4000. There are two different crystals with different space groups ($P2_1$ for Apo-1; $P4_1$ for Apo-2) appeared in the same crystallization condition drop. Crystallization conditions for SeMet KS were 0.8 M ammonium tartrate dibasic, pH 7.0. For the inhibitor BPH-629 complex, the KS protein solution containing 1 mM $MgCl_2$ and 10 mM DTT was co-crystallized with 2.5 mM BPH-629 under the same conditions as the wild-type protein. To obtain the KS-CPP structure, the D75C mutant crystal was prepared under the same crystallization conditions and then soaked with the cryoprotectant containing 5 mM CPP for 5 hours. All crystals were prepared at 4^{o}C for 12 hours and were then transferred to room temperature. They reached suitable size for X-ray diffraction in 2-3 days. The cryoprotectant was 1.0 M ammonium tartrate dibasic, pH 7.0, 10% w/v PEG4000 and 20% glycerol.

X-ray diffraction data sets were processed by using the HKL2000 program[33]. Prior to structure refinement, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor of model quality[34]. The multiple-wavelength anomalous diffraction (MAD) data sets of the selenium-containing derivatives were collected at wavelengths of 0.9789 Å (peak), 0.9790 Å (inflection) and 0.9611 Å (remote). Using SOLVE and RESOLVE[35], we obtained the figure of merit (FOM) values of 0.52, Z-scores from 82.58 to 90.28, and up to 430 auto-built amino acid residues. The BPH-629 and CPP complex structures were solved by using molecular replacement. All following structural refinements were carried out by using Coot[36] and CNS[37]. Data collection and refinement statistics are summarized in Supplementary Table 5.1, 5.3, 5.5 and 5.6. All graphics for the protein structures as well as electrostatic surface calculation were prepared by using the PyMOL program (http://www.pymol.org/).

BjKS Mutant Preparation

The KS mutants were created by using QuickChange Site-Directed Mutagenesis Kit with KS-pET32 Xa/LIC gene as the template. The mutagenic oligonucleotides were 5'-GCGCTGCTGTTCTGGCTCTGCGATTGCAACGACCTTGG-3' for D75C, 5'-CTGGCGCTGCTGTTCTGGCTCGCGGATTGCAACGACCTTGGCCTG-3' for D75A, 5'-TGGCTCGACGATTGCAACTGCCTTGGCCTGATCTCGCC-3' for D79C, and 5'-

CGGCTCATCTCCGCGATAGGGGCGCTGCAGAACGATCTGCATGGA-3' for R204A. The constructs were transformed to *E. coli* BL21trxB (DE3) for protein expression. The expression and purification of mutants were the same as those used for the wild-type enzyme.

SCORECONS Analysis of BJKS

A library of 10 KS sequence homologs and 342 AgABS sequence homologs were obtained by using the Jpred 3 server[24]. Multiple sequence alignment was carried out with the ClustalW2 server[38]. The aligned sequence library was then uploaded to the SCORECONS server[25] to calculate the conservation scores for each residue of BjKS.

Activity Assay for BjKS Proteins

BjKS activity assays were carried out as described previously[39]. Briefly, the dephosphorylation and cyclization of *ent*-CPP catalyzed by BjKS was monitored by using a continuous spectrophotometric assay for diphosphate release[40] in 96 well plates with 200 μL reaction mixtures containing 400 μM 2-amino-6-mercapto-7-methylpurine (MESG), 50 μM *ent*-CPP, 25 mM Tris-HCl (pH 7.5), 0.01% Triton X-100 and 1 mM $MgCl_2$. The IC_{50} values were obtained by fitting the inhibition data to a rectangular hyperbolic dose-response function in OriginPro 8.5 (OriginLab, Northampton, MA).

5.6 Schemes, Charts, Tables and Figures

Scheme 5.1 Reactions catalyzed by AgABS, PpCPPS/KS and BjKS, and proposed catalytic mechanism for the BjKS reaction. (a) Abietadiene is formed from GGPP by a bifunctional terpene cyclase abietadiene synthase, in *Abies grandis*. (b) Formation of *ent*-CPP, *ent*-kaurene and kauranol from GGPP catalyzed by *Physcomitrella patens* CPPS/KS. (c) Two proteins are used to produce *ent*-kaurene in the soil bacterium, *Bradyrhizodium japonicum*. Proposed carbocation intermediates are shown.

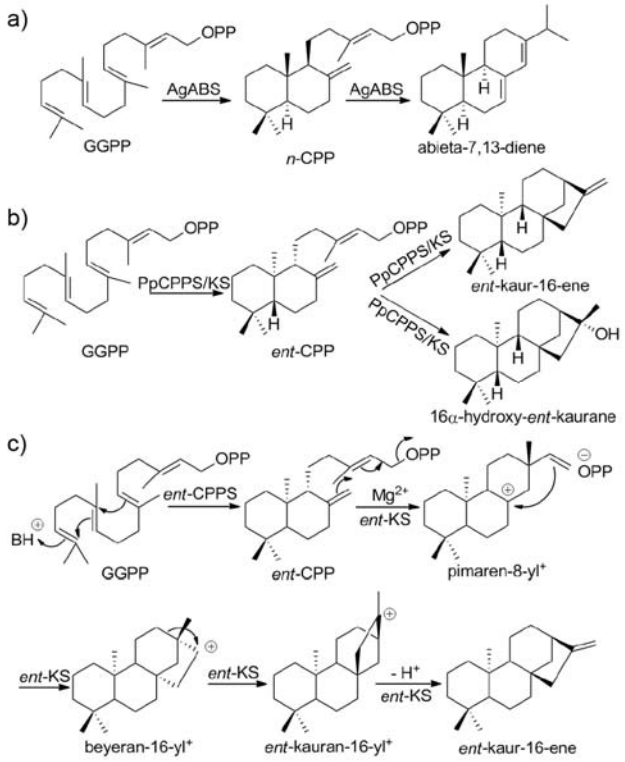


Table 5.1 Summary of Se-Met KS protein data collection and phasing statistics. Values in parentheses are for the outermost resolution shells.

Name	Se-Met KS		
	peak	inflection	remote
Data collection			
Wavelength (Å)	0.9789	0.9790	0.9611
Resolution (Å)	25.0 - 2.7 (2.8 - 2.7)	25.0 - 2.7 (2.8 - 2.7)	25.0 - 2.7 (2.8 - 2.7)
Space group	$P4_1$	$P4_1$	$P4_1$
Unit-cell *a*, *b*, *c* (Å)	66.0, 66.0, 136.3	66.0, 66.0, 136.3	66.0, 66.0 136.3
No. of unique reflections	31514 (3170)	31480 (3200)	31668 (3210)
Redundancy	4.1 (4.2)	4.2 (4.1)	4.1 (4.0)
Completeness (%)	99.6 (100.0)	99.8 (100.0)	99.8 (99.9)
Mean $I/\sigma(I)$	16.1 (3.9)	15.6 (3.8)	14.3 (3.0)
R_{merge} (%)	9.7 (38.8)	10.1 (39.8)	10.9 (48.4)
Phasing			
No. of sites		10	
Z-score		82.58 - 90.28	
Figure of merit		0.52	

Table 5.2 Data collection and refinement statistics for KS crystals. Values in parentheses are for the outermost resolution shells.

	Apo-1	Apo-2	D75C+CPP	WT+ BPH- 629
Data collection				
Resolution (Å)	25.0 – 2.0 (2.1 – 2.0)	25.0 - 1.9 (2.0 - 1.9)	25.0 – 1.8 (1.9 – 1.8)	25.0 - 2.0 (2.1 - 2.0)
Space group	$P2_1$	$P4_1$	$P4_1$	$P4_1$
Unit-cell *a, b, c* (Å)	65.0, 130.1, 66.4	65.8, 65.8, 135.2	65.8, 65.8, 136.2	65.8, 65.8, 135.7
β (°)	95.6			
No. of Unique reflections	72234 (6728)	43643 (4390)	51614 (5123)	38697 (3888)
Redundancy	3.0 (2.7)	5.6 (5.8)	4.9 (4.4)	3.4 (3.3)
Completeness (%)	98.0 (91.5)	99.8 (100.0)	99.5 (98.9)	99.4 (100.0)
Mean I/σ(I)	18.1 (3.1)	42.9 (4.5)	42.3 (3.6)	29.2 (3.3)
R_{merge} (%)	5.6 (20.7)	5.9 (46.7)	3.5 (37.2)	3.9 (46.3)
Refinement				
No. of chain(s)	4	2	2	2
No. of reflections	70362 (6036)	42679 (3982)	50837 (3310)	37623 (3499)
R_{work} (95% of data)	0.176 (0.230)	0.191 (0.251)	0.206 (0.28)	0.197 (0.279)
R_{free} (5% of data)	0.220 (0.276)	0.232 (0.287)	0.239 (0.309)	0.235 (0.291)
R.m.s.d. bonds (Å)	0.020	0.010	0.008	0.010
R.m.s.d. angles (°)	1.7	1.3	1.2	1.3
Ramachandran plot (%)				
Most favored (%)	97.3	96.7	97.5	96.6
Allowed (%)	2.2	2.8	2.1	2.8
Disallowed (%)	0.5	0.5	0.4	0.6
Mean B ($Å^2$) /atoms	27.8 / 9417	39.3 / 4926	42.2 / 4540	48.0 / 4692
PDB ID code	4KT9	3W3F	3WBV	3W3H

Table 5.3 Summary of observed protein residues and ligands

Name	PDB ID	Chain ID	Residues	Ligands
Apo-1	4KT9	A	3-290	-
		B	3-281	-
		C	2-290 (213-220 are missing)	-
		D	2-280 (211-222 are missing)	-
Apo-2	3W3F	A	1-281 (211-221 are missing)	-
		B	3-290 (211-222 are missing)	-
D75C + CPP	3WBV	A	1-275 (212-220 are missing)	CPP
		B	2-277 (147-150 and 211-221 are missing)	CPP
WT + BPH-629	3W3H	A	1-281 (211-221 are missing)	629
		B	3-290 (149-150 and 211-222 are missing)	-

Table 5.4 Q-score table for homology tree shown in Figure 5.4.

Q score	1DI1	5EAS	1N1B	2ONG	1PS1	2F7M	2Q58	3MAV	1YHK	2FOR	1UBV	2EWG	1RQI	1RTR	2E8T	2J1O
1DI1	1.000	0.171	0.155	0.175	0.399	0.195	0.174	0.165	0.173	0.134	0.178	0.160	0.175	0.161	0.114	0.151
5EAS	0.171	1.000	0.549	0.591	0.182	0.140	0.105	0.119	0.127	0.100	0.131	0.120	0.130	0.120	0.092	0.115
1N1B	0.155	0.549	1.000	0.722	0.174	0.116	0.086	0.100	0.109	0.099	0.104	0.089	0.112	0.102	0.083	0.098
2ONG	0.175	0.591	0.722	1.000	0.190	0.132	0.102	0.114	0.112	0.102	0.119	0.104	0.120	0.098	0.098	0.102
1PS1	0.399	0.182	0.174	0.190	1.000	0.180	0.157	0.158	0.174	0.140	0.182	0.162	0.173	0.169	0.110	0.149
2F7M	0.195	0.140	0.116	0.132	0.180	1.000	0.443	0.602	0.602	0.374	0.780	0.538	0.411	0.366	0.332	0.336
2Q58	0.174	0.105	0.086	0.102	0.157	0.443	1.000	0.490	0.532	0.335	0.403	0.491	0.388	0.337	0.290	0.299
3MAV	0.165	0.119	0.100	0.114	0.158	0.602	0.490	1.000	0.547	0.308	0.479	0.525	0.371	0.308	0.290	0.283
1YHK	0.173	0.127	0.109	0.112	0.174	0.602	0.532	0.547	1.000	0.379	0.524	0.805	0.459	0.354	0.324	0.302
2FOR	0.134	0.100	0.099	0.102	0.140	0.374	0.335	0.308	0.379	1.000	0.335	0.340	0.729	0.587	0.268	0.555
1UBV	0.178	0.131	0.104	0.119	0.182	0.780	0.403	0.479	0.524	0.335	1.000	0.484	0.378	0.364	0.312	0.326
2EWG	0.160	0.120	0.089	0.104	0.162	0.538	0.491	0.525	0.805	0.340	0.484	1.000	0.443	0.326	0.333	0.290
1RQI	0.175	0.130	0.112	0.120	0.173	0.411	0.388	0.371	0.459	0.729	0.378	0.443	1.000	0.559	0.326	0.530
1RTR	0.161	0.120	0.102	0.098	0.169	0.366	0.337	0.308	0.354	0.587	0.364	0.326	0.559	1.000	0.277	0.552
2E8T	0.114	0.092	0.083	0.098	0.110	0.332	0.290	0.290	0.324	0.268	0.312	0.333	0.326	0.277	1.000	0.233
2J1O	0.151	0.115	0.098	0.102	0.149	0.336	0.299	0.283	0.302	0.555	0.326	0.290	0.530	0.552	0.233	1.000
2Q80	0.155	0.109	0.092	0.116	0.151	0.397	0.388	0.351	0.379	0.346	0.364	0.391	0.429	0.362	0.551	0.314
1EZF	0.158	0.110	0.094	0.107	0.149	0.146	0.139	0.114	0.144	0.147	0.136	0.134	0.180	0.153	0.126	0.168
2ZCP	0.199	0.124	0.113	0.125	0.176	0.148	0.137	0.135	0.158	0.154	0.137	0.143	0.140	0.159	0.118	0.167
BjKS	0.244	0.190	0.184	0.198	0.307	0.211	0.170	0.171	0.199	0.181	0.216	0.207	0.191	0.199	0.139	0.170
3P5R	0.123	0.376	0.323	0.331	0.138	0.092	0.078	0.076	0.088	0.072	0.079	0.081	0.083	0.071	0.070	0.072
3S9V	0.116	0.376	0.325	0.339	0.130	0.090	0.074	0.074	0.086	0.069	0.085	0.078	0.093	0.073	0.066	0.085
3PYA	0.089	0.241	0.212	0.233	0.088	0.048	0.046	0.043	0.047	0.043	0.045	0.047	0.049	0.045	0.029	0.049
2J5C	0.158	0.473	0.588	0.592	0.173	0.104	0.082	0.091	0.095	0.090	0.098	0.086	0.108	0.100	0.066	0.095
3V1V	0.262	0.162	0.152	0.182	0.343	0.163	0.122	0.141	0.142	0.116	0.152	0.136	0.150	0.126	0.112	0.129
3LGK	0.323	0.174	0.168	0.178	0.386	0.168	0.159	0.150	0.165	0.152	0.163	0.162	0.191	0.167	0.142	0.152
4KT8	0.005	0.010	0.012	0.013	0.010	0.014	0.012	0.011	0.014	0.016	0.015	0.012	0.013	0.013	0.017	0.017
2E98	0.019	0.012	0.010	0.011	0.015	0.018	0.021	0.016	0.018	0.022	0.020	0.017	0.017	0.016	0.020	0.022
2VFW	0.019	0.012	0.011	0.012	0.013	0.018	0.016	0.016	0.015	0.020	0.019	0.014	0.020	0.019	0.028	0.026
2VG4	0.018	0.010	0.011	0.013	0.018	0.014	0.011	0.011	0.014	0.018	0.014	0.013	0.018	0.020	0.020	0.021
3KRF	0.089	0.067	0.061	0.062	0.085	0.217	0.192	0.194	0.217	0.350	0.221	0.205	0.371	0.395	0.137	0.429
1O6H	0.007	0.021	0.024	0.024	0.010	0.009	0.012	0.012	0.007	0.013	0.007	0.007	0.011	0.010	0.010	0.018
1W6K	0.015	0.023	0.025	0.025	0.010	0.011	0.012	0.008	0.008	0.013	0.010	0.010	0.014	0.010	0.008	0.013
3SDQ	0.099	0.359	0.360	0.375	0.116	0.084	0.065	0.070	0.077	0.071	0.072	0.070	0.082	0.079	0.056	0.079

Q score	2Q80	1EZF	2ZCP	KS	3P5R	3S9V	3PYA	2J5C	3V1V	3LGK	4KT8	2E98	2VFW	2VG4	3KRF	1O6H	1W6K	3SDQ
1DI1	0.155	0.158	0.199	0.244	0.123	0.116	0.089	0.158	0.262	0.323	0.005	0.019	0.019	0.018	0.089	0.007	0.015	0.099
5EAS	0.109	0.110	0.124	0.190	0.376	0.376	0.241	0.473	0.162	0.174	0.010	0.012	0.012	0.010	0.067	0.021	0.023	0.359
1N1B	0.092	0.094	0.113	0.184	0.323	0.325	0.212	0.588	0.152	0.168	0.012	0.010	0.011	0.011	0.061	0.024	0.025	0.360
2ONG	0.116	0.107	0.125	0.198	0.331	0.339	0.233	0.592	0.182	0.178	0.013	0.011	0.012	0.013	0.062	0.024	0.025	0.375
1PS1	0.151	0.149	0.176	0.307	0.138	0.130	0.088	0.173	0.343	0.386	0.010	0.015	0.013	0.018	0.085	0.010	0.010	0.116
2F7M	0.397	0.146	0.148	0.211	0.092	0.090	0.048	0.104	0.163	0.168	0.014	0.018	0.018	0.014	0.217	0.009	0.011	0.084
2Q58	0.388	0.139	0.137	0.170	0.078	0.074	0.046	0.082	0.122	0.159	0.012	0.021	0.016	0.011	0.192	0.012	0.012	0.065
3MAV	0.351	0.114	0.135	0.171	0.076	0.074	0.043	0.091	0.141	0.150	0.011	0.016	0.016	0.011	0.194	0.012	0.008	0.070
1YHK	0.379	0.144	0.158	0.199	0.088	0.086	0.047	0.095	0.142	0.165	0.014	0.018	0.015	0.014	0.217	0.007	0.008	0.077
2FOR	0.346	0.147	0.154	0.181	0.072	0.069	0.043	0.090	0.116	0.152	0.016	0.022	0.020	0.018	0.350	0.013	0.013	0.071
1UBV	0.364	0.136	0.137	0.216	0.079	0.085	0.045	0.098	0.152	0.163	0.015	0.020	0.019	0.014	0.221	0.007	0.010	0.072
2EWG	0.391	0.134	0.143	0.207	0.081	0.078	0.047	0.086	0.136	0.162	0.012	0.017	0.014	0.013	0.205	0.007	0.010	0.070
1RQI	0.429	0.180	0.140	0.191	0.083	0.093	0.049	0.108	0.150	0.191	0.013	0.017	0.020	0.018	0.371	0.011	0.014	0.082
1RTR	0.362	0.153	0.159	0.199	0.071	0.073	0.045	0.100	0.126	0.167	0.013	0.016	0.019	0.020	0.395	0.010	0.010	0.079
2E8T	0.551	0.126	0.118	0.139	0.070	0.066	0.029	0.066	0.112	0.142	0.017	0.020	0.028	0.020	0.137	0.010	0.008	0.056
2J1O	0.314	0.168	0.167	0.170	0.072	0.085	0.049	0.095	0.129	0.152	0.017	0.022	0.026	0.021	0.429	0.018	0.013	0.079
2Q80	1.000	0.177	0.185	0.168	0.089	0.082	0.041	0.095	0.144	0.191	0.021	0.028	0.026	0.020	0.206	0.011	0.010	0.071
1EZF	0.177	1.000	0.337	0.169	0.080	0.079	0.038	0.088	0.128	0.151	0.023	0.022	0.023	0.018	0.111	0.011	0.011	0.073
2ZCP	0.185	0.337	1.000	0.202	0.093	0.090	0.043	0.108	0.136	0.159	0.018	0.026	0.024	0.021	0.106	0.013	0.015	0.082
KS	0.168	0.169	0.202	1.000	0.145	0.147	0.071	0.179	0.262	0.301	0.016	0.028	0.024	0.026	0.119	0.017	0.016	0.133
3P5R	0.089	0.080	0.093	0.145	1.000	0.669	0.451	0.282	0.120	0.131	0.005	0.011	0.008	0.006	0.031	0.089	0.071	0.595
3S9V	0.082	0.079	0.090	0.147	0.669	1.000	0.533	0.271	0.113	0.128	0.007	0.004	0.009	0.009	0.046	0.087	0.069	0.577
3PYA	0.041	0.038	0.043	0.071	0.451	0.533	1.000	0.179	0.058	0.051	0.010	0.009	0.007	0.006	0.023	0.103	0.086	0.403
2J5C	0.095	0.088	0.108	0.179	0.282	0.271	0.179	1.000	0.154	0.176	0.010	0.009	0.012	0.012	0.055	0.021	0.022	0.298
3V1V	0.144	0.128	0.136	0.262	0.120	0.113	0.058	0.154	1.000	0.284	0.011	0.015	0.013	0.014	0.073	0.006	0.011	0.098
3LGK	0.191	0.151	0.159	0.301	0.131	0.128	0.051	0.176	0.284	1.000	0.010	0.020	0.014	0.016	0.090	0.006	0.008	0.119
4KT8	0.021	0.023	0.018	0.016	0.005	0.007	0.010	0.010	0.011	0.010	1.000	0.275	0.369	0.307	0.020	0.012	0.011	0.008
2E98	0.028	0.022	0.026	0.028	0.011	0.004	0.009	0.009	0.015	0.020	0.275	1.000	0.458	0.443	0.026	0.008	0.010	0.008
2VFW	0.026	0.023	0.024	0.024	0.008	0.009	0.007	0.012	0.013	0.014	0.369	0.458	1.000	0.552	0.027	0.017	0.010	0.010
2VG4	0.020	0.018	0.021	0.026	0.006	0.009	0.006	0.012	0.014	0.016	0.307	0.443	0.552	1.000	0.024	0.008	0.011	0.008
3KRF	0.206	0.111	0.106	0.119	0.031	0.046	0.023	0.055	0.073	0.090	0.020	0.026	0.027	0.024	1.000	0.013	0.014	0.051
1O6H	0.011	0.011	0.013	0.017	0.089	0.087	0.103	0.021	0.006	0.006	0.012	0.008	0.017	0.008	0.013	1.000	0.487	0.073
1W6K	0.010	0.011	0.015	0.016	0.071	0.069	0.086	0.022	0.011	0.008	0.011	0.010	0.010	0.011	0.014	0.487	1.000	0.058
3SDQ	0.071	0.073	0.082	0.133	0.595	0.577	0.403	0.298	0.098	0.119	0.008	0.008	0.010	0.008	0.051	0.073	0.058	1.000

Table 5.5 Q-score table for homology tree shown in Figure 5.3d.

Q score	1DI1	5EAS	1N1B	2ONG	1PS1	2F7M	2Q58	3MAV	1YHK	2FOR	1UBV	2EWG	1RQI	1RTR	2E8T	2J1O	2Q80	1EZF	2ZCP	KS	3P5R	3S9V	3PYA	2J5C	3V1V	3LGK	3KRF	3SDQ
1DI1	1.000	0.233	0.246	0.258	0.413	0.202	0.180	0.171	0.180	0.139	0.194	0.166	0.181	0.167	0.118	0.157	0.160	0.163	0.206	0.252	0.249	0.248	0.201	0.244	0.363	0.334	0.092	0.252
5EAS	0.233	1.000	0.621	0.621	0.274	0.207	0.194	0.215	0.184	0.189	0.201	0.173	0.207	0.195	0.195	0.175	0.230	0.171	0.168	0.324	0.631	0.615	0.305	0.668	0.352	0.270	0.127	0.551
1N1B	0.246	0.621	1.000	0.823	0.300	0.223	0.201	0.205	0.184	0.179	0.209	0.177	0.220	0.202	0.196	0.180	0.240	0.171	0.184	0.318	0.633	0.616	0.306	0.801	0.330	0.244	0.128	0.554
2ONG	0.258	0.621	0.823	1.000	0.303	0.215	0.192	0.210	0.198	0.172	0.208	0.186	0.206	0.208	0.199	0.176	0.243	0.181	0.202	0.316	0.611	0.597	0.314	0.771	0.351	0.291	0.128	0.548
1PS1	0.413	0.274	0.300	0.303	1.000	0.180	0.157	0.158	0.175	0.140	0.192	0.162	0.173	0.169	0.110	0.149	0.151	0.149	0.176	0.307	0.323	0.316	0.183	0.308	0.459	0.386	0.085	0.284
2F7M	0.202	0.207	0.223	0.215	0.180	1.000	0.443	0.602	0.605	0.374	0.823	0.538	0.411	0.366	0.332	0.336	0.397	0.146	0.148	0.211	0.189	0.230	0.120	0.198	0.218	0.168	0.217	0.207
2Q58	0.180	0.194	0.201	0.192	0.157	0.443	1.000	0.490	0.535	0.335	0.425	0.491	0.388	0.337	0.290	0.299	0.388	0.139	0.137	0.170	0.207	0.219	0.126	0.190	0.164	0.159	0.192	0.161
3MAV	0.171	0.215	0.205	0.210	0.158	0.602	0.490	1.000	0.550	0.308	0.505	0.525	0.371	0.308	0.290	0.283	0.351	0.114	0.135	0.171	0.217	0.211	0.123	0.206	0.189	0.150	0.194	0.173
1YHK	0.180	0.184	0.184	0.198	0.175	0.605	0.535	0.550	1.000	0.381	0.556	0.809	0.462	0.356	0.325	0.304	0.381	0.144	0.158	0.200	0.204	0.195	0.131	0.169	0.191	0.166	0.219	0.192
2FOR	0.139	0.189	0.179	0.172	0.140	0.374	0.335	0.308	0.381	1.000	0.353	0.340	0.729	0.587	0.268	0.555	0.346	0.147	0.154	0.181	0.189	0.202	0.117	0.180	0.155	0.152	0.350	0.174
1UBV	0.194	0.201	0.209	0.208	0.192	0.823	0.425	0.505	0.556	0.353	1.000	0.510	0.398	0.384	0.329	0.344	0.384	0.143	0.144	0.228	0.174	0.201	0.123	0.187	0.214	0.172	0.234	0.188
2EWG	0.166	0.173	0.177	0.186	0.162	0.538	0.491	0.525	0.809	0.340	0.510	1.000	0.443	0.326	0.333	0.290	0.391	0.134	0.143	0.207	0.186	0.176	0.123	0.160	0.182	0.162	0.205	0.173
1RQI	0.181	0.207	0.220	0.206	0.173	0.411	0.388	0.371	0.462	0.729	0.398	0.443	1.000	0.559	0.326	0.530	0.429	0.180	0.140	0.191	0.204	0.251	0.146	0.206	0.201	0.191	0.371	0.202
1RTR	0.167	0.195	0.202	0.208	0.169	0.366	0.337	0.308	0.356	0.587	0.384	0.326	0.559	1.000	0.277	0.552	0.362	0.153	0.159	0.199	0.207	0.208	0.135	0.199	0.168	0.167	0.395	0.194
2E8T	0.118	0.195	0.196	0.199	0.110	0.332	0.290	0.290	0.325	0.268	0.329	0.333	0.326	0.277	1.000	0.233	0.551	0.126	0.118	0.139	0.155	0.164	0.074	0.152	0.150	0.142	0.137	0.139
2J1O	0.157	0.175	0.180	0.176	0.149	0.336	0.299	0.283	0.304	0.555	0.344	0.290	0.530	0.552	0.233	1.000	0.314	0.168	0.167	0.170	0.163	0.180	0.113	0.165	0.172	0.152	0.429	0.195
2Q80	0.160	0.230	0.240	0.243	0.151	0.397	0.388	0.351	0.381	0.346	0.384	0.391	0.429	0.362	0.551	0.314	1.000	0.177	0.185	0.168	0.214	0.221	0.116	0.202	0.192	0.191	0.206	0.174
1EZF	0.163	0.171	0.171	0.181	0.149	0.146	0.139	0.114	0.144	0.147	0.143	0.134	0.180	0.153	0.126	0.168	0.177	1.000	0.337	0.169	0.164	0.170	0.223	0.167	0.171	0.151	0.111	0.179
2ZCP	0.206	0.168	0.184	0.202	0.176	0.148	0.137	0.135	0.158	0.154	0.144	0.143	0.140	0.159	0.118	0.167	0.185	0.337	1.000	0.202	0.191	0.219	0.103	0.188	0.182	0.159	0.106	0.201
3W7F	0.252	0.324	0.318	0.316	0.307	0.211	0.170	0.171	0.200	0.181	0.228	0.207	0.191	0.199	0.139	0.170	0.168	0.169	0.202	1.000	0.320	0.357	0.179	0.349	0.349	0.301	0.119	0.328
3P5R	0.249	0.631	0.633	0.611	0.323	0.189	0.207	0.217	0.204	0.189	0.174	0.186	0.204	0.207	0.155	0.163	0.214	0.164	0.191	0.320	1.000	0.720	0.341	0.589	0.332	0.303	0.120	0.635
3S9V	0.248	0.615	0.616	0.597	0.316	0.230	0.219	0.211	0.195	0.202	0.201	0.176	0.251	0.208	0.164	0.180	0.221	0.170	0.219	0.357	0.720	1.000	0.316	0.584	0.354	0.287	0.142	0.663
3PYA	0.201	0.305	0.306	0.314	0.183	0.120	0.126	0.123	0.131	0.117	0.123	0.123	0.146	0.135	0.074	0.113	0.116	0.223	0.103	0.179	0.341	0.316	1.000	0.283	0.210	0.166	0.095	0.293
2J5C	0.244	0.668	0.801	0.771	0.308	0.198	0.190	0.206	0.169	0.180	0.187	0.160	0.206	0.199	0.152	0.165	0.202	0.167	0.188	0.349	0.589	0.584	0.283	1.000	0.362	0.289	0.110	0.559
3V1V	0.363	0.352	0.330	0.351	0.459	0.218	0.164	0.189	0.191	0.155	0.214	0.182	0.201	0.168	0.150	0.172	0.192	0.171	0.182	0.349	0.332	0.354	0.210	0.362	1.000	0.380	0.098	0.321
3LGK	0.334	0.270	0.244	0.291	0.386	0.168	0.159	0.150	0.166	0.152	0.172	0.162	0.191	0.167	0.142	0.152	0.191	0.151	0.159	0.301	0.303	0.287	0.166	0.289	0.380	1.000	0.090	0.293
3KRF	0.092	0.127	0.128	0.128	0.085	0.217	0.192	0.194	0.219	0.350	0.234	0.205	0.371	0.395	0.137	0.429	0.206	0.111	0.106	0.119	0.120	0.142	0.095	0.110	0.098	0.090	1.000	0.125
3SDQ	0.252	0.551	0.554	0.548	0.284	0.207	0.161	0.173	0.192	0.174	0.188	0.173	0.202	0.194	0.139	0.195	0.174	0.179	0.201	0.328	0.635	0.663	0.293	0.559	0.321	0.293	0.125	1.000

Table 5.6. Activities of BjKS wild-type and mutant proteins using *ent*-CPP substrate and a PPi-release assay.

protein	specific activity, nmol $min^{-1}mg^{-1}$
WT	64±2.9
D75C	2.4±0.32
D75A	0.81±0.18
D79C	0.20±0.86
R204A	0.77±0.51
CPP only	-0.32±0.21

Figure 5.1 Structure of BjKS. (a) Stereoview of (Apo-1) BjKS (PDB ID code 4KT9) showing highly α-helical structure. (b) Dimer interface in *B. japonicum ent*-kaurene synthase. There is a cluster of side-chains of hydrophobic residues (M235, Y232 and F240, cyan spheres) from both chains at the "bottom" of the interface which form a hydrophobic core; further up, the E244 side-chain interacts with the N-terminus of the helix starting at Y155 in the other subunit (green spheres), and salt bridges exist between the E157(A) and R250(B) side-chains, and between E157(B) and R250(A) (purple spheres). The two predicted catalytic motifs (DDXXD and NTE) are also shown with spheres (red and orange, respectively). (c) Gel filtration results show BjKS exists as a dimer in solution (Mr = 61.0 kDa).

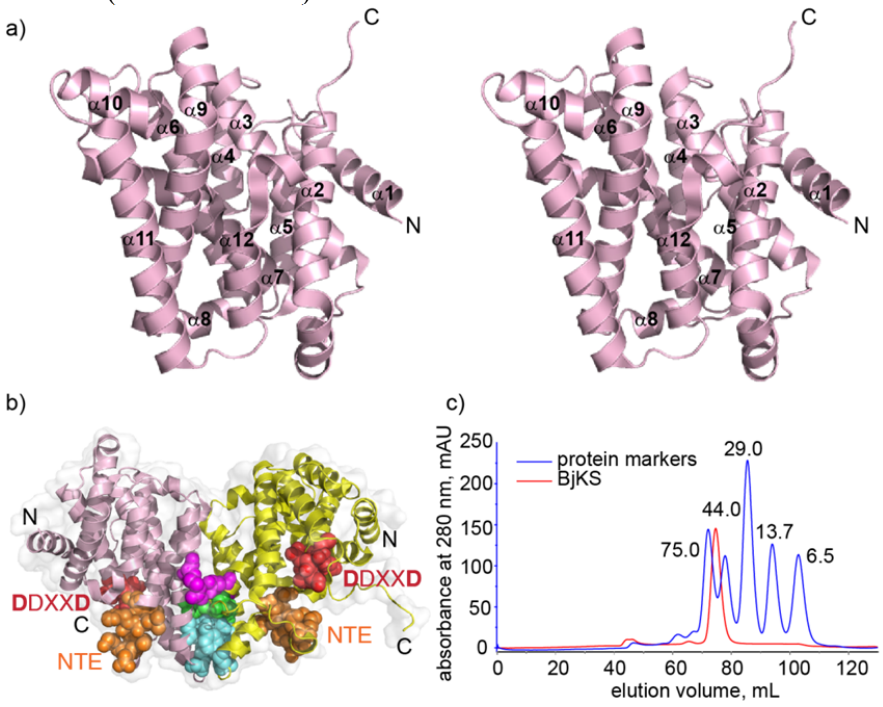


Figure 5.2 Gel filtration calibration. (a) Protein makers used. (b) Calibration curve for protein markers.

a)

protein	Log M_r	K_{av}	V_e, ml	Concentration, mg/ml
conalbumin	1.88	0.33	72.11	0.8
ovalbumin	1.63	0.41	77.84	1
carbonic anhydrase	1.46	0.52	85.49	0.8
ribonuclease A	1.14	0.63	93.95	0.8
aprotinin	0.81	0.76	102.91	0.8

b)

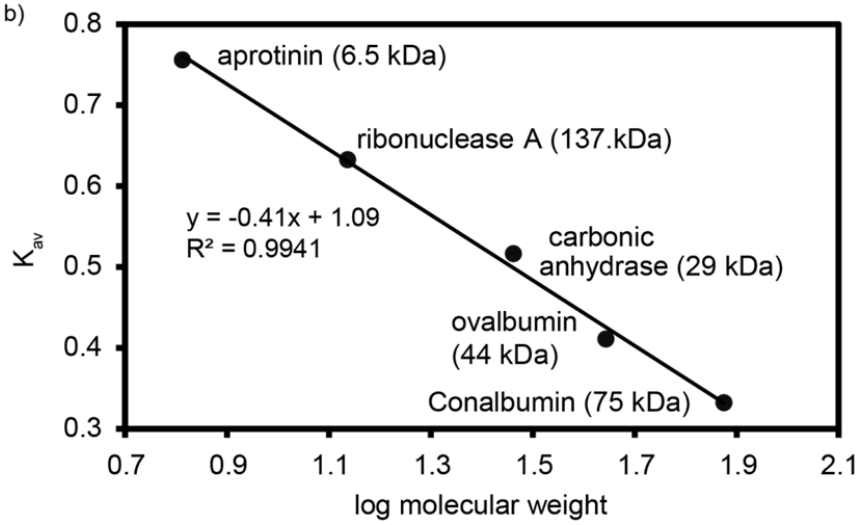


Figure 5.3 Superimposition of the overall structure of BjKS (red; PDB ID 4KT9) with other terpene cyclases. (a) Superimposition of BjKS with pentalenene synthase from *Streptomyces* UC5319 which has only an α domain (green; 1HM7); (b) Superimposition of BjKS with 1, 8-cineole synthase from *Salvia fruticosa* which has two domains (αβ; blue; 2J5C); (c) Superimposition of BjKS with abietadiene synthase from *Abies grandis* which has three domains (αβγ; yellow; 3S9V); d) Homology tree (based on α domain Q-scores[22, 23]) showing clustering of BjKS with the plant α domain proteins. Note that the α domains in the terpene cyclase (top) are distant from the α domains in the synthases such as FPPS, and have previously been referred to as "α_c" domains[19].

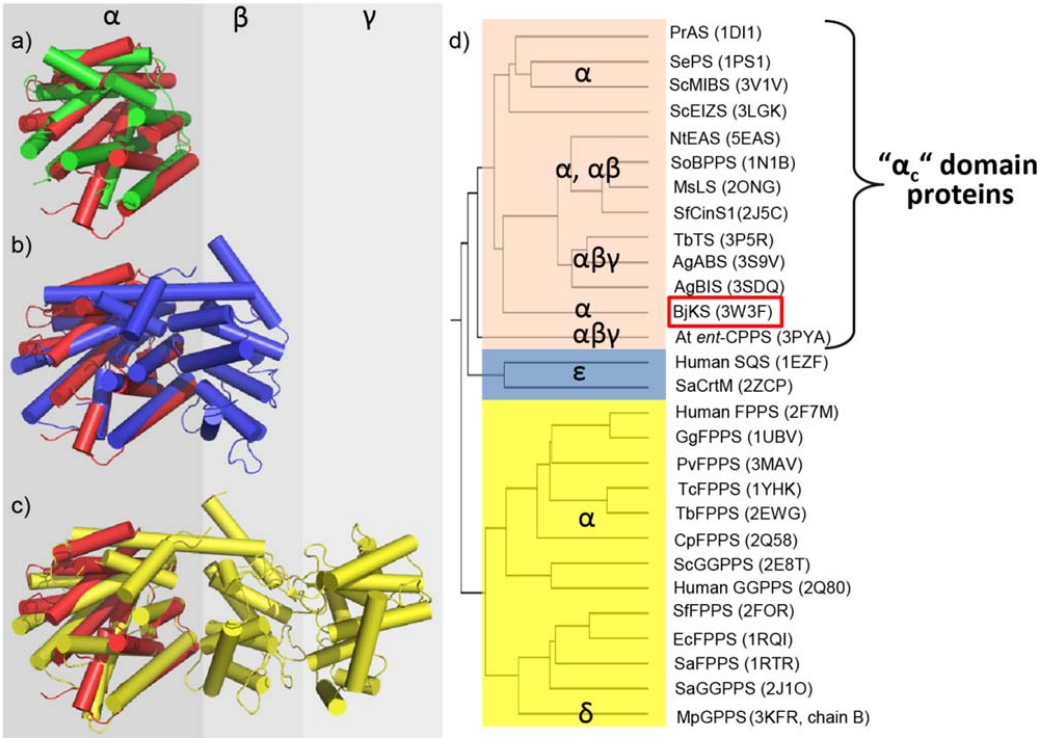


Figure 5.4 Homology tree of α, β, γ, δ, ε, and ζ-fold proteins obtained by using the Q-score program.

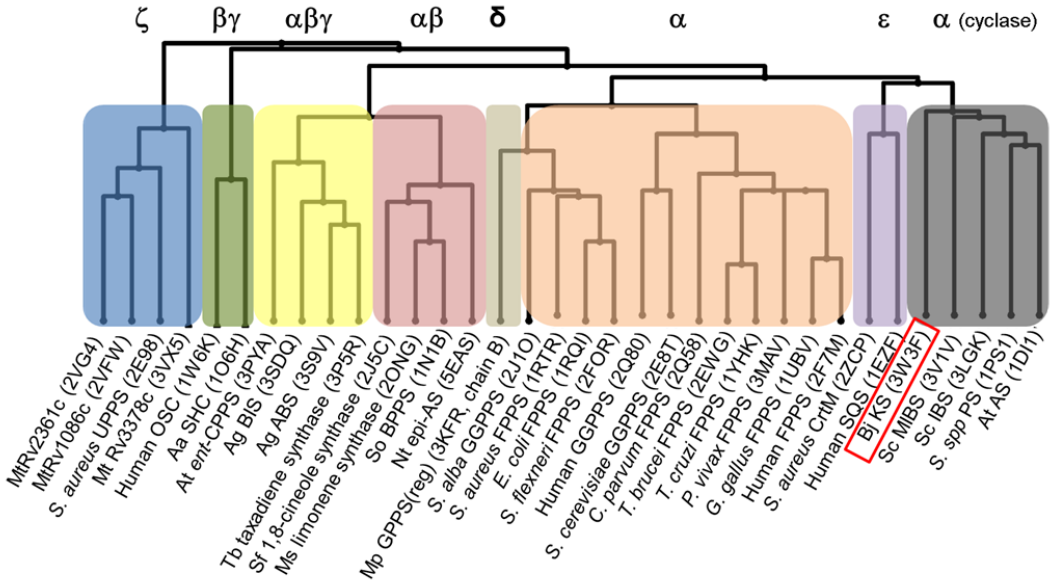
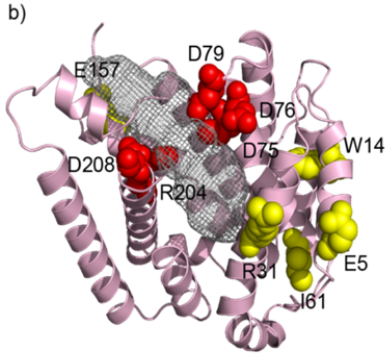


Figure 5.5 Essential residues and catalysis. (a) SCORECONS[25] results. (b) Highly conserved residues; residues in red were mutated to A or C. (c) Activity of mutant proteins.

a)

Ranking	Score	Residue
1	0.806	W14
2	0.78	I61
3	0.75	D79
4	0.741	R31
5	0.735	D76
6	0.731	D75
7	0.718	E5
8	0.714	R204
9	0.699	D208
10	0.684	E157

b)



c)

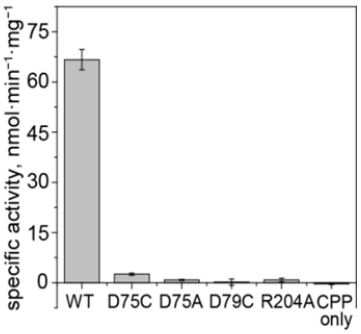


Figure 5.6 Sequence alignment of *Bradyrhizobium japonicum ent*-kaurene synthase (BjKS) with abietadiene synthase (AgABS), taxadiene synthase (TbTXS), 1, 8-cineole synthase (SfCinS1), *ent*-kaurene synthase (AtKS) and (+)-bornyl diphosphate synthase (BPPS). Sequence alignments were performed with ClustalX[41]. Strictly conserved residues are highlighted in red, conservatively substituted residues are boxed. The conserved DDXXD and RL(N,D)DXX(S,T,G)XXX(E,D) motifs are boxed. Mg^{2+}-binding residues are indicated by an asterisk. The Figure was produced by using ESPript[42].

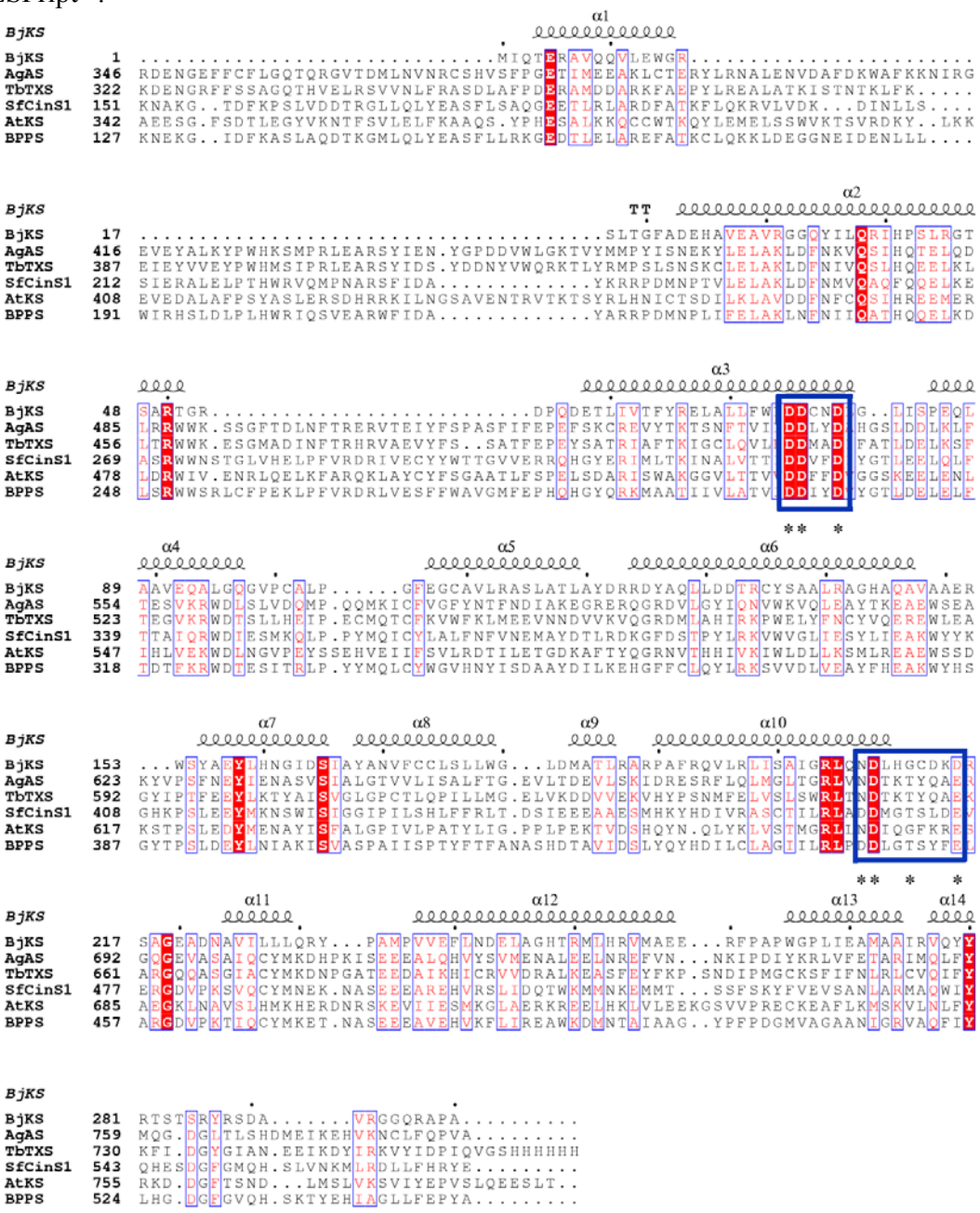


Figure 5.7 Ligand electron density results contoured at 1σ. (a) *ent*-CPP in chain A (PDB ID 3WBV). (b) *ent*-CPP in chain B (PDB ID 3WBV). (c) BPH-629 (PDB ID 3W3H).

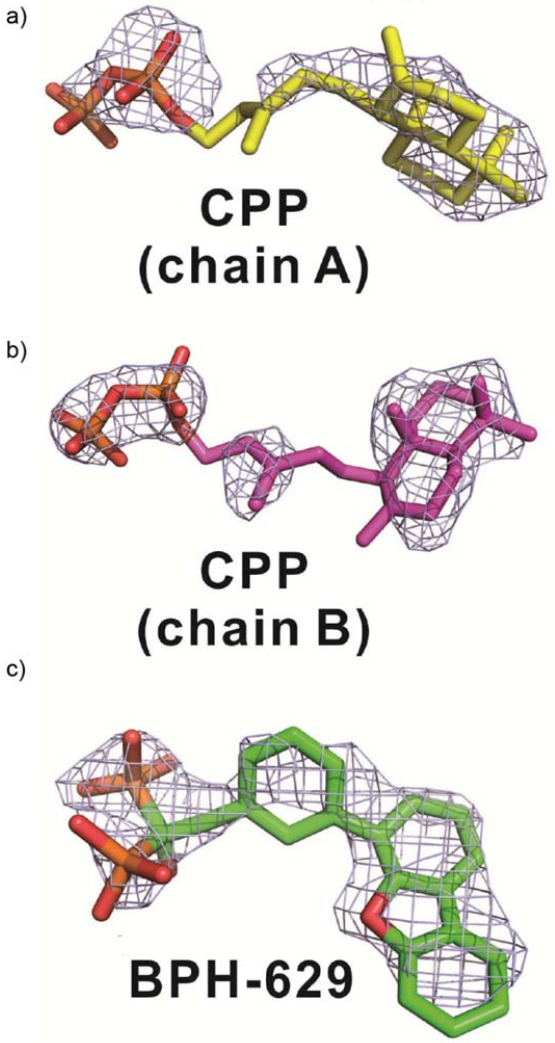


Figure 5.8 BjKS *ent*-CPPS structure. (a) Stereo-view of dimer structure (PDB ID 3WBV). (b) Rotated stereo-view of structure of BjKS in a). (c) Stereo-view of structure of *ent*-CPP (yellow) bound to BjKS superimposed on 2-F-geranylgeranyl diphosphate (FGG, cyan) bound to the α domain of taxadiene synthase (TXS) (PDB ID code 3P5R). The highly conserved "DDXXD" Asps of BjKS are indicated. The 3 Mg^{2+} are from the TXS structure and are not observed in BjKS. The hydrophobic side-chain of *ent*-CPP is buried in a hydrophobic pocket composed of I36, L68, L71, F72, Y136, Y168 and N207, shown as an orange surface.

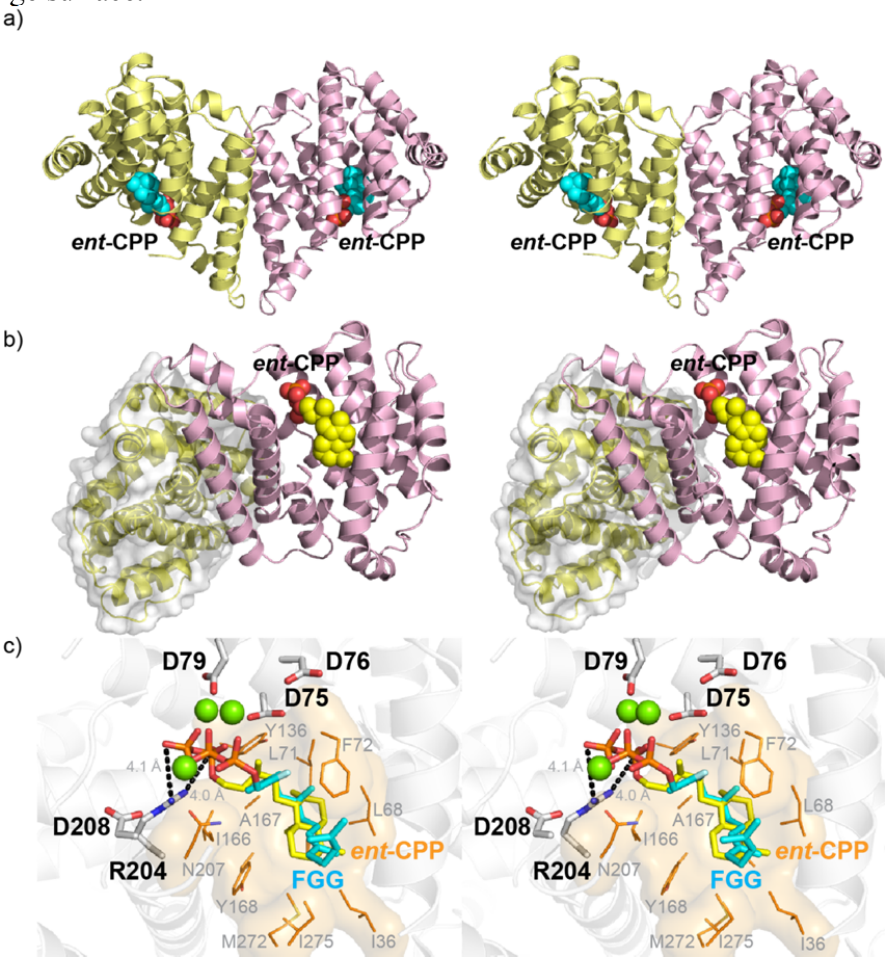


Figure 5.9 BPH-629 inhibition and inhibitor-bound structure. (a) IC_{50} for BPH-629 inhibition. (b) Possible hydrogen bonds form between BPH-629 and BjKS activity site residues in BPH-629/BjKS structure (PDB ID 3W3H). (c) Stereo view of structure of BjKS bound to the bisphosphonate inhibitor BPH-629 (pink) superimposed on *ent*-CPP (yellow).

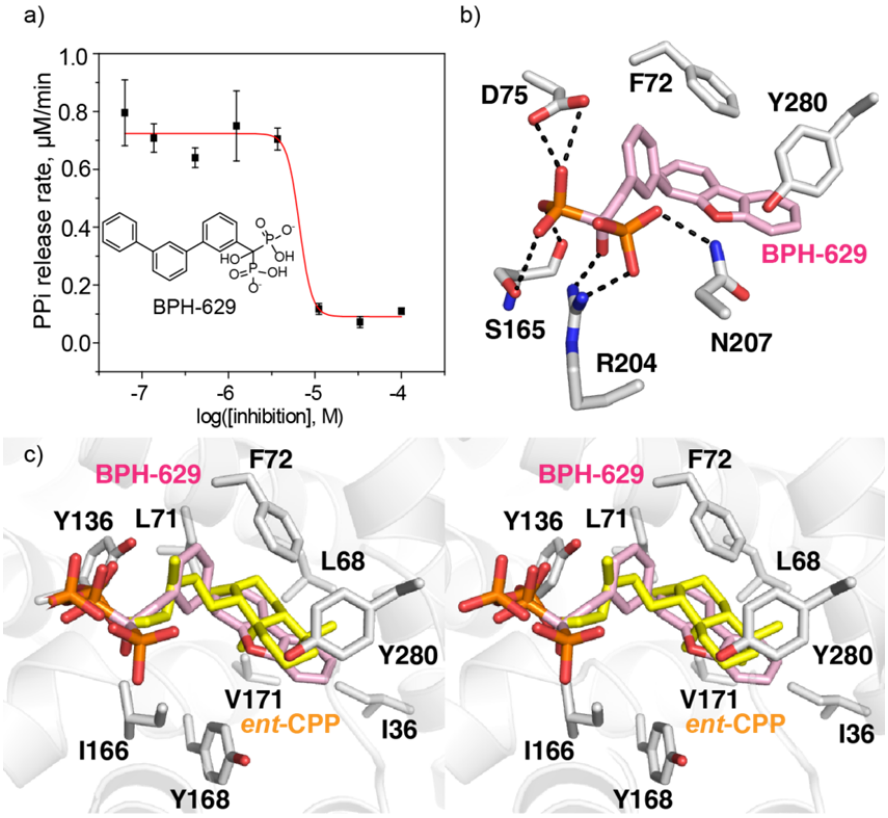
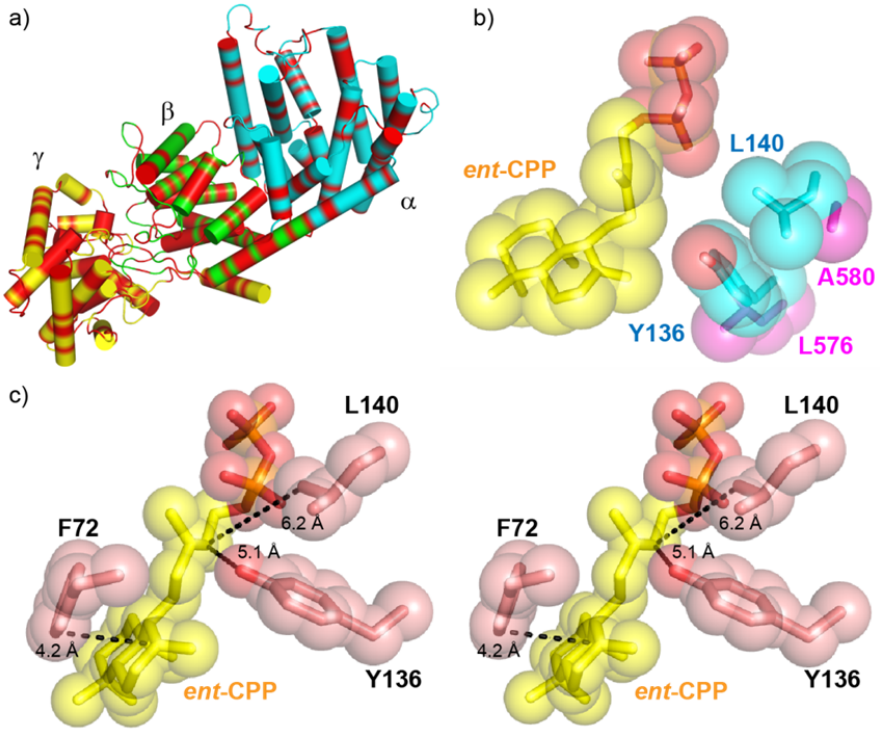


Figure 5.10 (a) Phyre2[28] homology model of *Physcomitrella patens ent*-CPPS/*ent*-KS showing the 45% of residues (in red) that are identical to those in abietadiene synthase, uniformly dispersed over the α, β and γ domains. (b) *ent*-CPP (yellow) bound to BjKS (PDB ID 3WBV) superimposed on corresponding *P. patens* KS residues. The large BjKS residues adjacent *ent*-CPP are in cyan. The smaller *P. patens* residues are in magenta and permit water quenching of the kauranyl-16-yl cation. (c) Stereo view of residues that are proposed to stabilize the carbocation intermediates in BjKS (PDB ID 3WBV). Distances between *ent*-CPP C12 and Y136 O, *ent*-CPP C12 and L140 C^{δ}, *ent*-CPP C8 and F72 C are 5.1 Å, 6.2 Å and 4.2 Å, respectively.



5.7 References

(1) Buckingham, J. *CRC, Boca Raton, FL* 2007.
(2) Oldfield, E.; Lin, F. Y. *Angew. Chem. Int. Ed .Engl*. 2012, *51*, 1124.
(3) Dairi, T. *J. Antibiot. (Tokyo).* 2005, *58*, 227.
(4) Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J. *FEBS Lett.* 2009, *583*, 475.
(5) Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H.; Dairi, T. *J. Biol. Chem.* 2002, *277*, 37098.
(6) Smanski, M. J.; Peterson, R. M.; Shen, B. In *Method Enzymol.*; David, A. H., Ed.; Academic Press: 2012; Vol. 515, p 163.
(7) Komakr, H.; Nemoto, A.; Tanaka, Y.; Takgi, H.; Yazawa, K.; Mikami, Y.; Shigemori, H.; Kobayashi, J.; Akikazu, A.; Nagata, Y. *J. Antibiot. (Tokyo).* 1999, *52*, 13.
(8) Tully, R. E.; van Berkum, P.; Lovins, K. W.; Keister, D. L. *BBA-Gene Struct. Expr.* 1998, *1398*, 243.
(9) Smanski, M. J.; Peterson, R. M.; Huang, S.-X.; Shen, B. *Curr. Opin. Chem. Biol.* 2012, *16*, 132.
(10) Mann, F. M.; Prisic, S.; Hu, H.; Xu, M.; Coates, R. M.; Peters, R. J. *J. Biol. Chem.* 2009, *284*, 23574.
(11) Prach, L.; Kirby, J.; Keasling, J. D.; Alber, T. *FEBS J.* 2010, *277*, 3588.
(12) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem. Int. Ed.* 2000, *39*, 2812.
(13) Wendt, K. U.; Schulz, G. E. *Structure* 1998, *6*, 127.
(14) Koksal, M.; Hu, H.; Coates, R. M.; Peters, R. J.; Christianson, D. W. *Nat. Chem. Biol.* 2011, *7*, 431.
(15) Kawaide, H.; Hayashi, K.-i.; Kawanabe, R.; Sakigi, Y.; Matsuo, A.; Natsume, M.; Nozaki, H. *FEBS J.* 2011, *278*, 123.
(16) Von Schwartzenberg, K.; Schultze, W.; Kassner, H. *Plant Cell Rep.* 2004, *22*, 780.
(17) MacMillan, J.; Beale, M. H. *Diterpene Biosynthesis*; Pergamon Press, 2000.
(18) Zhou, K.; Gao, Y.; Hoy, J. A.; Mann, F. M.; Honzatko, R. B.; Peters, R. J. *J. Biol. Chem.* 2012, *287*, 6840.
(19) Cao, R.; Zhang, Y.; Mann, F. M.; Huang, C.; Mukkamala, D.; Hudock, M. P.; Mead, M. E.; Prisic, S.; Wang, K.; Lin, F.-Y.; Chang, T.-K.; Peters, R. J.; Odfield, E. *Proteins* 2010, *78*, 2417.
(20) Chang, T.-H.; Hsieh, F.-L.; Ko, T.-P.; Teng, K.-H.; Liang, P.-H.; Wang, A. H. J. *Plant Cell* 2010, *22*, 454.
(21) Lesburg, C. A.; Zhai, G. Z.; Cane, D. E.; Christianson, D. W. *Science* 1997, *277*, 1820.
(22) Krissinel, E.; Henrick, K. *Acta Crystallogr. D.* 2004, *60*, 2256.
(23) Krissinel, E.; Henrick, K. Protein structure comparison service Fold at European Bioinformatics Institute. http://www.ebi.ac.uk/msd-srv/ssm.
(24) Cole, C.; Barber, J. D.; Barton, G. J. *Nucleic Acids Res.* 2008, *36*, W197.
(25) Valdar, W. S. *Proteins* 2002, *48*, 227.
(26) Koeksal, M.; Hu, H.; Coates, R. M.; Peters, R. J.; Christianson, D. W. *Nat. Chem. Biol.* 2011, *7*, 431.
(27) McAndrew, R. P.; Peralta-Yahya, P. P.; DeGiovanni, A.; Pereira, J. H.; Hadi, M. Z.; Keesling, J. D.; Adams, P. D. *Structure* 2011, *19*, 1876.

(28) Kelley, L. A.; Sternberg, M. J. E. *Nat. Protoc.* 2009, *4*, 363.
(29) Christianson, D. W. *Chem. Rev.* 2006, *106*, 3412.
(30) Dougherty, D. A. *Science* 1996, *271*, 163.
(31) Jenson, C.; Jorgensen, W. L. *J. Am. Chem. Soc.* 1997, *119*, 10846.
(32) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *J. Mol. Biol.* 1993, *229*, 105.
(33) Otwinowski, Z.; Minor, W. *Method Enzymol., Pt A* 1997, *276*, 307.
(34) Brunger, A. T. *Acta Crystallogr. D.* 1993, *49*, 24.
(35) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D.* 2010, *66*, 213.
(36) Emsley, P.; Cowtan, K. *Acta Crystallogr. D.* 2004, *60*, 2126.
(37) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D.* 1998, *54*, 905.
(38) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* 2007, *23*, 2947.
(39) Durrant, J. D.; Cao, R.; Gorfe, A. A.; Zhu, W.; Li, J.; Sankovsky, A.; Oldfield, E.; McCammon, J. A. *Chem. Biol. Drug. Des.* 2011, *78*, 323.
(40) Webb, M. R. *Proc. Natl. Acad. Sci. U. S. A.* 1992, *89*, 4884.
(41) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. *Nucleic Acids Res.* 1997, *25*, 4876.
(42) Gouet, P.; Robert, X.; Courcelle, E. *Nucleic Acids Res.* 2003, *31*, 3320.

Chapter 6: Structure and Inhibition of Tuberculosinol Synthase and Decaprenyl Diphosphate Synthase from *Mycobacterium tuberculosis*

6.1 Notes and Acknowledgements

H.-C Chan, X. Feng, C.-H. Huang, Y. Hu, W. Liu and Y. Zheng, cloned, expressed and purified Rv3378c proteins. X. Feng and S. Bogue cloned, expressed and purified Rv2361c proteins. H.-C Chan, T.-P. Ko and R.-T. Guo carried out crystallization, data collection, structure determination and refinement of Rv3378c proteins. X. Feng and S. Bogue carried out crystallization, data collection, structure determination and refinement of Rv3378c proteins. X. Feng and S. Bogue characterized the enzymatic activity of Rv3378c and Rv2361c, and carried out mutagenesis study. X. Feng and S. Bogue carried out ITC studies. C. Nakano and T. Hoshino provide tuberculosinyl diphosphate, and measured Rv3378c mutants activities. X. Feng, P.-H. Liang, A. H.-J. Wang, E. Oldfield and R.-T. Guo analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by the National Basic Research Program of China (2011CB710800 and 2011CBA00805), the Tianjin Municipal Science and Technology Commission (12ZCZDSY12500), and the United States Public Health Service (National Institutes of Health grant GM065307). X.F. was supported by a Pre-doctoral Fellowship from the American Heart Association, Midwest Affiliate (13PRE14510056). We thank the National Synchrotron Radiation Research Center of Taiwan for beamtime allocation and data collection assistance.

This chapter was reproduced in part with permission from H.-C Chan, X. Feng, T.-P. Ko, C.-H. Huang, Y. Hu, Y. Zheng, S. Bogue, C. Nakano, T. Hoshino, L. Zhang, P. Lv, W. Liu, D. C. Crick, P.-H. Liang, A. H.-J. Wang, E. Oldfield, R.-T. Guo. *J. Am. Chem. Soc.*, 2014, 136 (7), pp 2892–2896. H.-C Chan, X. Feng. and T.-P. Ko contributed equally to this work.

6.2 Introduction

Tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, is a major source of morbidity and mortality worldwide with almost two million deaths annually, and the rise in multi-drug resistant and extensively drug resistant strains is of great concern[1-3].

There is thus considerable interest in the development of new drugs and new leads such as SQ109 and TMC-207[4], which target cell wall biosynthesis and ATP formation, respectively, and also in the development of new therapeutic approaches that target virulence factor (VF) formation. VFs are by definition not essential for bacterial growth outside of host cells, but are involved in processes such as invasion, persistence, lysis, and evasion of innate immune system responses, as found for example with staphyloxanthin in *Staphylococcus aureus*[5,6]. This carotenoid VF provides resistance to host-based reactive oxygen species-based killing, and inhibiting staphyloxanthin biosynthesis is a novel route to anti-infective therapy[6]. In *M. tuberculosis*, one class of VFs are the tuberculosinols[7-9]: tuberculosinol (TOH) and the (±) *iso*-tuberculosinols (*iso*-TOH), Scheme 6.1. The latter were first isolated from a marine sponge from the Nosy Be island, Madagascar, and are thus also known as (±) nosyberkols[10] (Scheme 6.1). The original structures proposed[7-10] have been confirmed by total synthesis[11,12]. The biosynthesis of the tuberculosinols is catalyzed by two enzymes: Rv3377c, tuberculosinyl (halama-5,13-dien-15-yl) diphosphate synthase, and Rv3378c, tuberculosinol/(±)*iso*-tuberculosinol synthase (Scheme 6.1). Both proteins are essential for bacterial survival inside macrophages[13] with the tuberculosinols inhibiting phagolysosome maturation, as well as macrophage phagocytosis[8,9,14]. The three-dimensional structures of both Rv3377c and Rv3378c are thus of considerable interest since they are thought to represent novel anti-virulence therapeutic targets for tuberculosis[15]. Rv3377c is a diterpene cyclase that converts geranylgeranyl diphosphate (GGPP) into tuberculosinyl diphosphate (TPP), while Rv3378c converts TPP into TOH and the *iso*-TOHs, acting as a phosphatase/isomerase, as shown in Scheme 6.1. Here, we report the structure of Rv3378c both in its apo form as well as bound to its TPP substrate, and to a bisphosphonate inhibitor, BPH-629 (Scheme 6.1). We also report the structure of the *M. tuberculosis cis*-decaprenyl diphosphate synthase (Rv2361c) with another bisphosphonate inhibitor, BPH-640 (Scheme 6.1) which has a similar structure and inhibitor-binding mode as found with Rv3378c-the phosphatase.

6.3 Results and Discussions

We cloned, expressed, purified, crystallized and solved the Rv3378c protein structure by using multiple isomorphous replacement in a tetragonal crystal, then by molecular replacement in a monoclinic crystal, Table 6.1 and Table 6.2 in the Supporting Information. Residues 46-50 could not be modeled in one crystal, or 82-90 in another, due to lattice packing and disorder. Nevertheless, the two structures superimposed well and a continuous protein dimer model could be constructed (Figure 6.1a and Figure 6.2a in the Supporting Information), consistent with the observation that Rv3378c also exists as a dimer in solution[9]. The overall fold shows close structural homology to that seen in *M. tuberculosis cis*-farnesyl diphosphate synthase (Rv1086; 2.33 Å Cα rmsd over 201 residues; PDB ID code 2VFW), *M. tuberculosis* decaprenyl diphosphate synthase (Rv2361; 2.23 Å over 195 residues; PDB ID code 2VG4), and *E. coli* undecaprenyl diphosphate synthase (UPPS; 2.44 Å over 203 residues; PDB ID code 1JP3), although there is only ~12-14% sequence identity with these proteins.

There is a potential substrate-binding site at the "top" of the structure (Figure 6.1b and Figure 6.2b in the Supporting Information) but this cavity[16] is smaller than that seen in UPPS, which makes the C_{55} diphosphate used in bacterial cell wall biosynthesis (~600 $Å^3$ *versus* ~900-1700 $Å^3$, depending on the presence or absence of bound ligands, Figure 6.1c and Figure 6.2c in the Supporting Information). The Rv3378c protein thus adopts the same fold as seen in *cis*-head-to-tail prenyl diphosphate synthases and this structure is unprecedented in enzymes involved in terpene biosynthesis. Attempts to obtain the structures of Rv3378c in the presence of its substrate TPP (by soaking and co-crystallization) were unsuccessful, so we next used computational docking[17] to see how TPP might bind to Rv3378c, to guide the design of inactive mutants. The docking results obtained (Figure 6.3 in the Supporting Information) indicated that the fused rings bound to a large hydrophobic pocket while the allyl side-chain docked close to Y51 and Y90, suggesting that these residues might be involved in activating water molecules for nucleophilic attack at C13 and/or C15. To test this hypothesis, we made a Y51F mutant and a Y51F/Y90F double mutant finding ~13% and ~5% wild type activity, respectively, Figure 6.4 in the Supporting Information. We then obtained structures of both mutants in the presence of TPP. In both, TPP occupies the polar-top/nonpolar-bottom cavity shown in Figure 6.1b (Figure 6.5a and Figure 6.6a in the Supporting Information), with more

complete electron densities found in the Y51F/Y90F double mutant (Figure 6.5b and Figure 6.6b in the Supporting Information, yellow) than in the Y51F single mutant (Figure 6.7 in the Supporting Information), due presumably to more residual activity in the Y51F single mutant crystal. A similar though not identical hydrophobic pocket is occupied by S-*thiolo* farnesyl diphosphate (FSPP) when bound to UPPS, as illustrated in the superposition shown in Figure 6.5b (FSPP in cyan). In UPPS, the two most conserved residues from a SCORECONS[18] analysis (which ranks residues in terms of their conserved nature) are D26 and R30 (Table 6.3 in the Supporting Information) which in UPPS are involved in binding to Mg^{2+} and farnesyl diphosphate (FPP), facilitating diphosphate activation and removal. These two residues correspond to D34 and R38 in Rv3378c (a 2.2 Å Cα rmsd), and based on this homology to UPPS, might therefore be expected to play a role in diphosphate activation and release.

To examine the role of these residues in catalysis, we mutated both to Ala. Activity results for these mutants together with Y51F, Y51F/Y90F, Y90F as well as the single DDXXD mutants reported previously[9] are shown in Figures 6.5c and 6.5d as well as in Figure 6.4 in the Supporting Information. Whereas the WT protein has 41% *iso*-TOH activity and 59% TOH activity, D34A, R38A and Y90F mutants produced on average ~1% *iso*-TOH but on average 49% TOH, meaning that D34, R38 and Y90 are all important for *iso*-TOH production, a more S_N1-like reaction that requires D34 and R38 to facilitate dephosphorylation and carbocation formation, Figure 6.5d. On the other hand, Y51 is involved in both TOH and *iso*-TOH formation since production of both is reduced in the Y51F mutant. In addition, the Y51F/Y90F double mutant has essentially no activity. Taken together, these results indicate that D34, R38, Y51 and Y90 are involved in 13R, 13S-*iso*-tuberculosinol formation (Figure 6.5d), while it is primarily Y51 that is involved in TOH formation, since both Y51F and the Y51F/Y90F mutants produce little or no TOH, as illustrated in Figure 6.5c. The single DDXXD mutants have activities ranging from 32% to 53%, whereas the double Asp mutants are not active[9]. This suggests that the DDXXD motif, although not immediately adjacent to the active site, may also play a role in catalysis, perhaps by helping remove the Mg^{2+}-PPi formed, since 2 Asps would be required for Mg^{2+} chelation.

Since Rv3378c is a target for anti-virulence therapeutics, we were also interested in discovering Rv3378c inhibitors. We identified BPH-629 as one such inhibitor having a K_d~560 nM, Figure 6.8a in the Supporting Information. Co-crystallization of Rv3378c and BPH-629 in the presence of Mg^{2+} yielded a complex structure containing two molecules of BPH-629 and one Mg^{2+} per Rv3378c monomer. Figure 6.9a and Figure 6.10a in the Supporting Information show the location of the inhibitor binding sites in the dimer. In each monomer, there are two bound BPH-629 molecules: one (BPH-629 (1)) is buried in the substrate binding site (Figure 6.9b and Figure 6.10b in the Supporting Information), the other (BPH-629 (2)) is located in the dimer interface, between a helix of monomer A and a loop of monomer B (Figures 6.9a and 6.9b). In both cases, the bisphosphonate head-groups chelate to a single Mg^{2+}, that is, in turn, coordinated to the side-chain of D34, Figure 6.9c and Figure 6.10c in the Supporting Information. Moreover, as shown in Figure 6.9c, the side-chain in BPH-629(1) makes extensive nonpolar interactions with numerous hydrophobic amino acids in the TPP binding site, and BPH-629(2) is sandwiched between H41 of monomer A and Y259, S260 and L267 of monomer B.

Interestingly, a similar surface binding site is also found in *M. tuberculosis* DPPS (Rv2361c) in complex with another bisphosphonate, BPH-640, a close analog of BPH-629, as shown in Figures 6.9d to 6.9f and Figures 6.11a to 6.11c in the Supporting Information. BPH-640 is a 410 nM inhibitor of DPPS (Figure 6.8b in the Supporting Information), but lacks the O and CH_2 groups present in BPH-629. Full crystallographic data acquisition and structure refinement details for this structure are shown in Table 6.1 and electron density results are in Figure 6.9f. As with BPH-629 (2), BPH-640 also occupies a dimer interface binding site, sandwiched between G77, N78, G79, R80, T83, R89 and R127 of monomer A and R292 and F293 of monomer B (Figure 6.9f).

The results shown in Figure 6.9 indicate that the terpene synthase (phosphatase) Rv3378c and the *cis*-prenyl transferase DPPS both have the same overall fold. They both also exist as dimers in solution[9,19] as well as in crystals (Figure 6.9a and 3d), and both are inhibited by structurally similar bisphosphonates, with closely-related binding modes (Figure 6.9b and 3e) and similar affinities. These observations lead to the idea that in the future it may be possible to design multi-target inhibitors[20-22] that target both enzymes in

TB bacteria inside macrophages, inhibiting growth (DPPS/cell wall biosynthesis) as well as virulence (Rv3378c/tuberculosinol biosynthesis/invasion).

6.4 Conclusions

We report the first structure of any bacterial diterpene synthase, tuberculosinol/*iso*-tuberculosinol synthase (Rv3378c) from *M. tuberculosis*, a target for anti-infective therapies that block virulence factor formation. This phosphatase adopts the same fold as found in the Z or *cis*-prenyltransferases. We also obtained structures in the presence of TPP and with a bisphosphonate inhibitor where one molecule occupied the TPP substrate binding site; the second located at the dimer interface site. An unusual surface site was also found with a related bisphosphonate inhibitor bound to the *M. tuberculosis* DPPS (Rv2361c). In addition, we used the X-ray structures of Rv3378c together with the results of site-directed mutagenesis to propose mechanisms of action for formation of tuberculosinol and the *iso*-tuberculosinols in which two Tyr residues play an important role. Given the similarity in local and global structure between Rv3378c and Rv2361,the possibility exists that in the future it might be possible to develop multi-target inhibitors that target not only virulence, but also cell wall biosynthesis, based in part on the structures reported here.

6.5 Materials and Methods

Cloning, expression and purification of Rv3378c (tuberculosinol/(13*R,S*)*iso*-tuberculosinol synthase)

The gene encoding Rv3378c from *Mycobacterium tuberculosis* H37Rv was synthesized chemically and amplified by polymerase chain reaction (PCR) with forward primer 5'-GGTATTGAGGGTCGCATGAACCTGGTTAGCGAGAAAGAGTTC-3' and reverse primer 5'-AGAGGAGAGTTAGAGCCATTAACCCTCAGCAAACCAGATACCGTCATG-3', then cloned into the pET32 Xa/LIC vector. The Rv3378c recombinant plasmids were transformed to *E. coli* BL21 (DE3) and protein expression induced with 0.4 mM isopropyl-thiogalactopyranoside (IPTG) at 20 °C for 24 hours. The cell lysate was harvested by centrifugation and the resulting supernatant loaded onto a Ni-NTA column

equilibrated in buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl, and 20 mM imidazole. His-tagged Rv3378c was eluted using an imidazole gradient (20-250 mM). The protein solution was then dialyzed against a buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl and subjected to Factor Xa digestion, to remove the thioredoxin and His tags. The mixture was then again passed through a Ni-NTA column, and untagged Rv3378c eluted with a 20 mM imidazole-containing buffer. Protein was further purified by FPLC using a DEAE column, 25 mM Tris, pH 7.5, and a 0-500 mM NaCl gradient. The protein eluted at ~225 mM NaCl and was then concentrated to 5 mg/mL in 25 mM Tris, pH 7.5, and 150 mM NaCl.

Crystallization, data collection and structure determination of Rv3378c

For tetragonal crystals, wild-type Rv3378c protein containing 10 mM $MgCl_2$ and 10 mM DTT was first crystallized by using the Crystal Screen I Kit (Hampton Research, Aliso Viejo, CA) and the sitting-drop vapor diffusion method. The reservoir solution (No. 48) contained 0.1 M Tris-HCl, pH 8.5, and 2.0 M $NH_4H_2PO_4$. Better crystals were obtained by optimizing the reservoir composition to 0.1 M Tris-HCl, pH 8.5, and 2.0-2.2 M $NH_4H_2PO_4$. All crystals were prepared at room temperature. They reached suitable size for X-ray diffraction in 2 days. For heavy-atom derivatives, the Hg-containing reagents of Heavy Atom Screen Hg (Hampton Research) were used. Cryoprotectant solutions (0.1 M Tris-HCl, pH 8.5, 2.2 M $NH_4H_2PO_4$, 2.5 mM $MgCl_2$ and 14% glycerol) containing 2 mM Hg derivatives were used in soaking the native crystals for 3 to 5 hours. For the BPH-629 inhibitor complex, the Rv3378c protein solution containing 2.5 mM $MgCl_2$ and 10 mM DTT was co-crystallized with 2.5 mM BPH-629 under the same conditions as described above. To obtain the Rv3378c-TPP structure, Y51F and Y51F/Y90F mutant crystals were prepared under the same crystallization conditions. Cryoprotectant solution (0.1 M Tris-HCl, pH 8.5, 2.2 M $NH_4H_2PO_4$, 2.5 mM $MgCl_2$ and 14% glycerol) containing 5 mM TPP was used in soaking the Y51F and Y51F/Y90F crystals, for 5 hours. For monoclinic native crystals, the wild-type Rv3378c protein containing 2.5 mM $MgCl_2$ and 10 mM DTT was crystallized in 0.15 M DL-malic acid, pH 7.0 and 19% PEG 3350. The cryoprotectant was 0.15 M DL-malic acid, pH 7.0, 19% PEG 3350, 14% glycerol, and 2.5 mM $MgCl_2$.

X-ray diffraction data sets were processed by using the HKL2000 program[23]. Prior to structure refinement, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor of model quality[24]. For Rv3378c, the MIR datasets of the mercury-containing derivatives were collected at a wavelength of 0.97622 Å. Using SOLVE and RESOLVE[25], combinations of datasets from different Hg-derivative crystals with the "native" dataset from the wild-type crystal improved the figure of merit (FOM) values from 0.52 to 0.66, the Z-scores from 10.85 to 14.1, and the number of auto-built amino acid residues to 181. Structures of the monoclinic crystals and the complex with BPH-629 were solved by molecular replacement using CNS[26] and a monomeric model from the tetragonal crystal. All subsequent structure refinements were carried out by using Coot[27] and CNS[26]. Data collection and refinement statistics are summarized in Table 6.1 and Table 6.2 in the Supporting Information. All Rv3378c crystals diffracted well to 2.1-2.8 A with R_{merge} of 0.050-0.065 (Table 6.1).

Crystallization, data collection and structure determination for Rv2361c (decaprenyl diphosphate synthase)

Rv2361c was expressed and purified as described previously[19]. For apo crystals, Rv2361c in 20 mM Tris (pH 7.2) containing 0.15 mM $MgCl_2$ and 3mM DTT was concentrated to 10mg/mL, and screened for crystallization using the JCSG Core Suites (QIAGEN, Valencia, CA) and the hanging drop method. Once initial crystallization conditions were found, they were optimized by varying the precipitant and protein concentrations. Optimal crystallization conditions were found to be 0.1 M HEPES (pH 7.5), 10% glycerol, 25% PEG 400, 7% PEG 3000, with 10 mg/mL Rv2361c. Apo Rv2361c crystals were then soaked in 1 mM BPH-640 in the same buffer for 5 hours to obtain inhibitor bound crystals. The X-ray diffraction data were indexed, integrated, and scaled by using the HKL2000 package[23]. Structures were determined by using molecular replacement with the program ccp4i[28] with the Rv2361c apo structure (PDB code 2VG4) as a template. Further model building and refinement employed Coot[27] and the ProDRG[29] server. All protein structure graphics were prepared by using the PyMOL program (http://www.pymol.org/).

Rv2361c Inhibition assay

Inhibition of Rv2361c by BPH-629 and BPH-640 was evaluated using [^{3}H]-IPP. Protein (0.1 μM) in 25 mM Tris buffer (pH 7.5) containing 0.01% Triton X100, 1mM $MgCl_2$, 15 μM GPP and compound in a total volume of 50 μL was incubated at room temperature for 10 min. Then, 3.6 μL of IPP (containing 0.9 nmol cold IPP and 54.24 nCi [^{3}H]-IPP) was added, and the resulting mixture incubated at 37 °C for 20 min. The reaction was stopped with 500 μM NaCl saturated H_2O. Prenyl alcohols were then extracted with 500 μL butanol, and a 300-μL aliquot was transferred to a scintillation vial for radioactivity counting. IC_{50} values were obtained by using Origin software (OriginLab, Northampton, MA).

Tuberculosinol/(13R,S) *iso*-tuberculosinol synthase activity assay

A reaction mixture (2.5 mL) containing 50 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 50 μg (44.4 μM) TPP, and 150 μg (1.76 μM) enzyme was incubated at 37°C for 14 hr. Reaction product analysis protocols were as described previously[30].

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) measurements were carried out as described previously[31].

6.6 Schemes, Charts, Tables and Figures

Scheme 6.1 Biosynthesis of *Mycobacterium tuberculosis* virulence factors (tuberculosinols) and decaprenyl diphosphate, which is essential for cell wall biosynthesis.

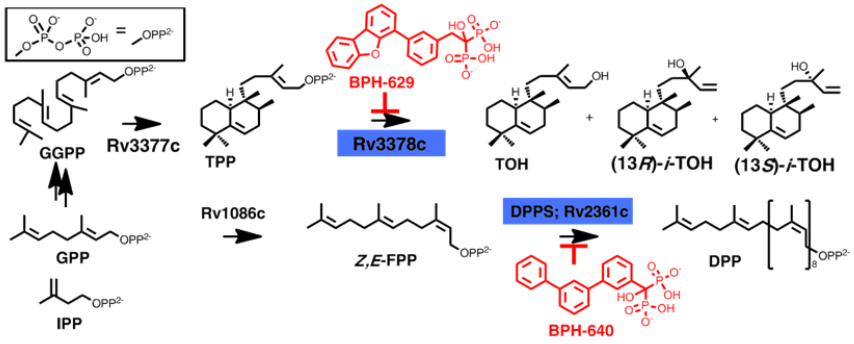


Table 6.1 Data collection and refinement statistics for the Rv3378c crystals. Values in parentheses are for the outermost resolution shells.

	Rv3378c Native 1	Rv3378c Native 2	Rv3378c Y51F with TPP	Rv3378c Y51F/Y90F with TPP	Rv3378c with 629	Rv2361c with 640
Data collection						
Resolution (Å)	25.0 - 2.3 (2.4 - 2.3)	25.0 - 2.1 (2.2 - 2.1)	25.0 - 2.7 (2.8 - 2.7)	25.0 – 2.4 (2.5 – 2.4)	25.0 - 2.1 (2.2 - 2.1)	50-1.8 (1.9-1.8)
Space group	$P4_32_12$	$C2$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P2_12_12_1$
Unit-cell *a* (Å)	105.9	208.4	106.5	106.3	105.4	77.8
b (Å)	105.9	56.5	106.5	106.3	105.4	89.2
c (Å)	66.6	114.4	66.4	66.8	66.0	93.9
β (°)	90.3	90.3	90.3	90.3	90.3	90.0
No. of Unique reflections	18334 (1794)	77301 (6681)	11190 (1049)	15419 (1505)	21574 (2082)	58370 (2875)
Redundancy	7.9 (8.1)	3.9 (2.7)	7.2 (7.7)	10.3 (10.6)	10.2 (10.2)	14.8 (14.8)
Completeness (%)	99.9 (100.0)	98.1 (85.6)	99.3 (96.8)	99.7 (100.0)	99.3 (97.6)	100.0 (100.0)
Mean I/σ(I)	43.6 (6.5)	19.2 (2.3)	47.1 (3.6)	37.2 (4.3)	52.4 (6.4)	41.6 (3.4)
R_{merge} (%)	5.0 (43.3)	6.5 (41.6)	5.6 (39.1)	6.5 (52.1)	5.3 (39.7)	7.8 (69.4)
Refinement						
No. of chain(s)	1	4	1	1	1	2
No. of reflections	17676 (1634)	65562 (2237)	10756 (946)	14914 (1326)	20950 (1961)	55357 (3783)
R_{work} (95% of data)	0.212 (0.250)	0.198 (0.222)	0.240 (0.342)	0.231 (0.293)	0.197 (0.242)	0.157 (0.227)
R_{free} (5% of data)	0.239 (0.308)	0.239 (0.302)	0.261 (0.365)	0.260 (0.363)	0.237 (0.243)	0.198 (0.294)
R.m.s.d. bonds (Å)	0.007	0.006	0.006	0.008	0.008	0.025
R.m.s.d. angles (°)	1.3	1.2	1.3	1.2	1.4	2.4
Ramachandran plot (%)						
Most favored (%)	95.1	96.3	94.4	94.7	96.9	94.0
Allowed (%)	4.9	3.7	5.6	5.3	3.1	5.8
Disallowed (%)	0.0	0.0	0.0	0.0	0.0	0.2
Mean B ($Å^2$) /atoms	55.2 / 2611	28.6 / 10274	89.0 / 2520	55.7 / 2444	46.8 / 2757	26.9 / 4510
PDB ID code	*3VX5*	*3VXA*	*3W3I*	*4KT8*	*3VX9*	*4H21*

Table 6.2 Summary of data collection and phasing statistics for Rv3378c crystals using Hg-derivatives. Values in parentheses are for the outer-most resolution shells.

Crystal	Native	$C_{13}H_{17}HgNO_6$	$(C_2H_5HgO)HPO_2$	$C(HgOOCCH_3)_4$	$C_7H_5ClHgO_2$
Data collection					
Resolution (Å)	25.0 - 2.3 (2.4 - 2.3)	25.0 - 2.3 (2.4 - 2.3)	25.0 - 2.3 (2.4 - 2.3)	25.0 - 2.4 (2.5 - 2.4)	25.0 - 2.3 (2.4 - 2.3)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell *a, b, c* (Å)	105.9, 105.9, 66.6	106.2, 106.2, 66.4	105.9, 105.9, 66.6	106.4, 106.4, 66.9	105.8, 105.8, 66.3
No. of Unique reflections	17923 (1750)	17431 (1707)	17392 (1687)	15685 (1542)	17194 (1676)
Redundancy	13.5 (14.4)	8.9 (9.1)	10.8 (11.0)	14.0 (14.5)	14.0 (14.4)
Completeness (%)	99.0 (100.0)	99.9 (100.0)	99.9 (100)	99.9 (100.0)	100.0 (99.8)
Mean $I/\sigma(I)$	52.5 (5.3)	30.1 (5.8)	47.4 (6.4)	42.6 (8.5)	60.9 (6.1)
R_{merge} (%)	6.2 (45.7)	4.9 (47.7)	7.6 (52.1)	7.6 (47.1)	6.1 (47.8)
Phasing					
No. of sites			6		
Z-score			14.1		
Figure of merit			0.66		

Table 6.3 SCORECONS result for EcUPPS (and homologous/aligned residues in Rv3378c).

Ranking	*E. coli* Residue No.	SCORECONS Score	Residue	*E. coli* UPPS Catalytic Role	Aligned Residue in Rv3378c
1	26	0.988	D	binds Mg^{2+}	D34
2	30	0.984	R	binds FPP	R38
3	28	0.976	N	H-bond FPP O1	T36
4	74	0.968	N	binds IPP	L84
5	20	0.964	H		V28
6	190	0.96	D		D202
7	77	0.953	R	binds FPP	G87
8	204	0.945	F		D214
9	145	0.941	Y		F161
10	194	0.937	R	binds IPP	G206

Figure 6.1 Structures of tuberculosinol/(13R,S)-iso-tuberculosinol synthase (Rv3378c) and undecaprenyl diphosphate synthase (UPPS). a) Rv3378c dimer. Both N and C termini of both monomers locate at the top of the dimer structure. b) Predicted Rv3378c ligand binding site. c) E. coli UPPS ligand binding sites (PDB ID code 2E98). Electrostatic surfaces are shown for Rv3378c and UPPS cavities, colored blue in positive region, red in negative region, calculated using PyMOL.

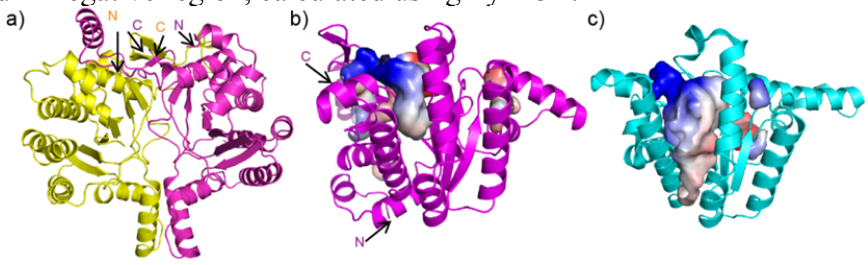


Figure 6.2 Stereo-views of the structures shown in Figure 6.1.

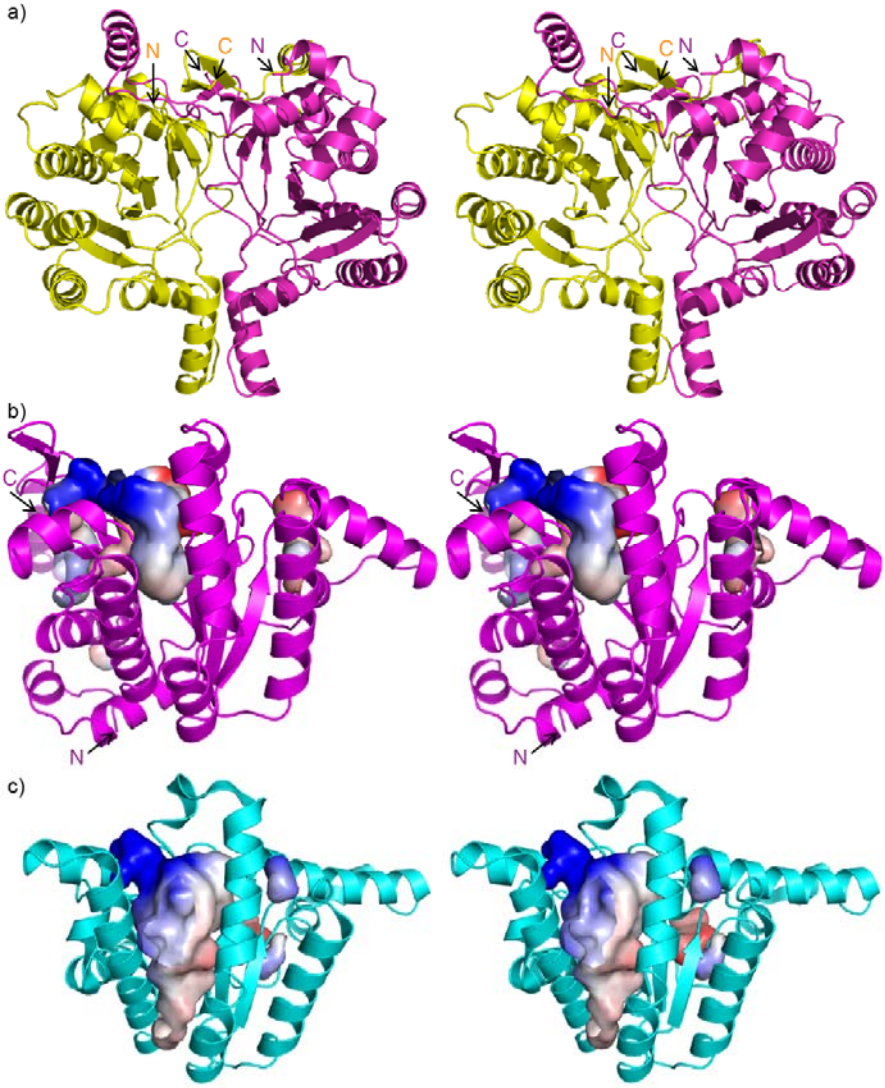


Figure 6.3 Docking pose for TPP in Rv3378c.

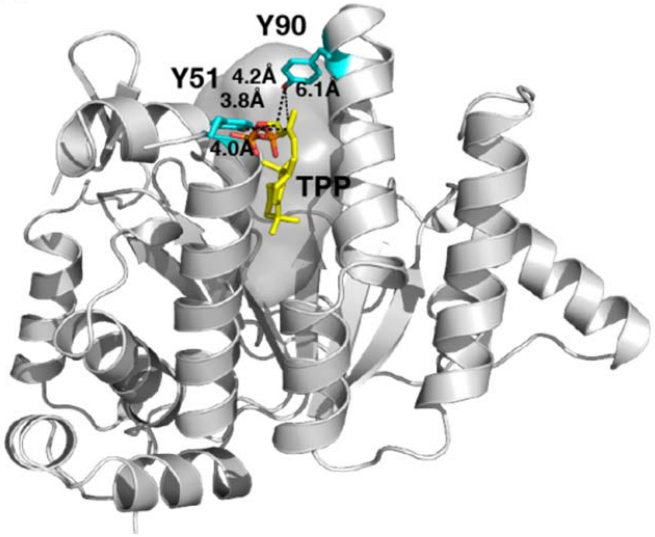


Figure 6.4 GC-MS chromatograms of mutant Rv3378c products.

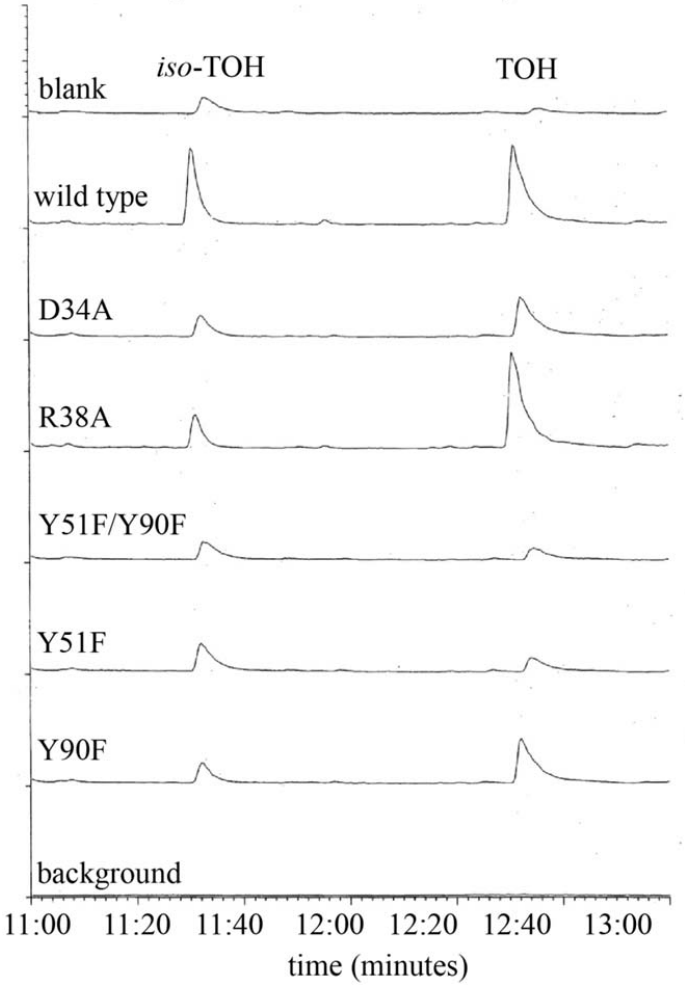


Figure 6.5 Structures and activities of Rv3378c. a) Rv3378c Y51F/Y90F double mutant structure with bound TPP. b) Rv3378c/TPP (yellow, electron density contoured at 1.5σ) structure superimposed on S-thiolo-FSPP (cyan), bound to UPPS (PDB code 1X06). The Mg^{2+} is from the UPPS structure. Catalytic residues in Rv3378c are shown with grey sticks. Y51, Y90 are built based on the F51, F90 coordinates. c) Activities of wild type and Rv3378c mutants in TOH production and proposed catalytic mechanism for TOH formation. d) Activities of wild type and Rv3378c mutants in iso-TOH production and proposed catalytic mechanism for iso-TOH formation.

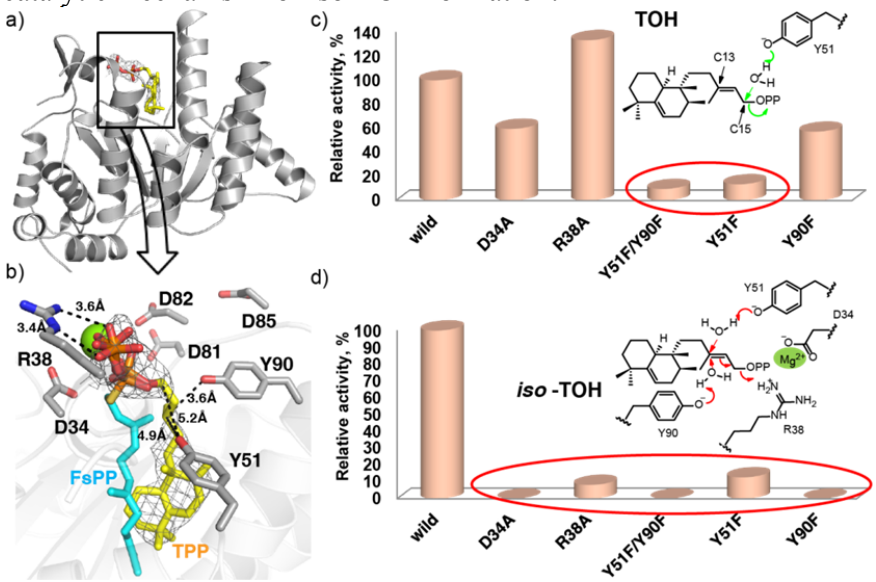


Figure 6.6 Stereo-views of the structures shown in Figure 6.5.

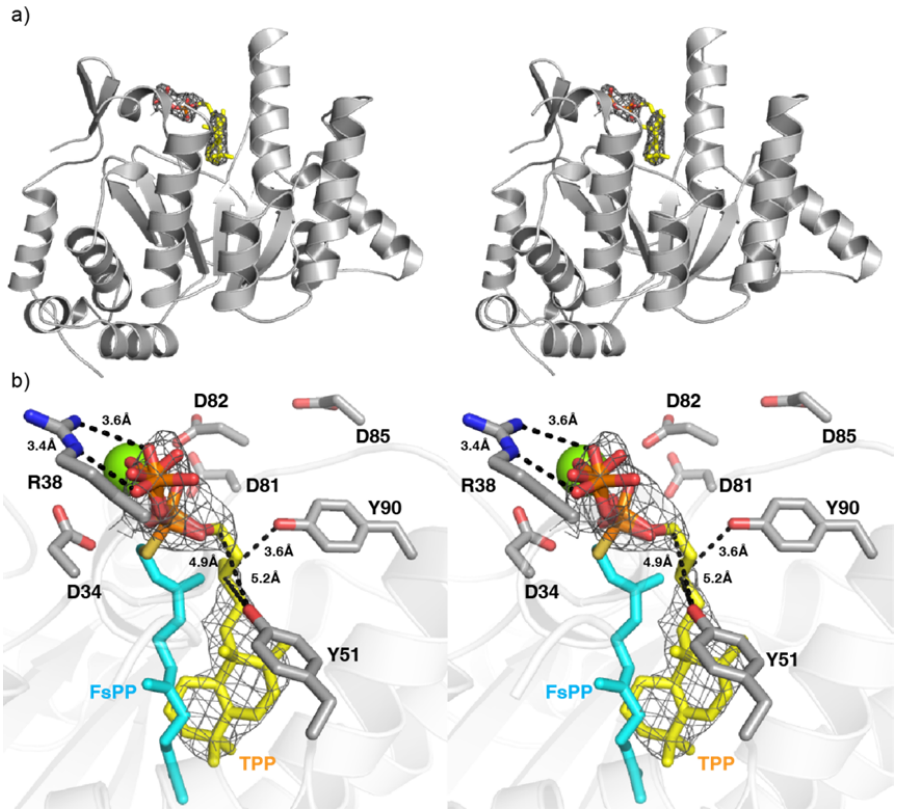


Figure 6.7 Stereo-view of the Rv3378c/Y51F-TPP complex structure, with TPP electron densities contoured at 1.5σ.

Figure 6.8 Isothermal titration calorimetry results for binding of bisphosphonates to Rv3378c, DPPS and FPPS[32]. a) ITC results for BPH-629 binding to Rv3378c. ΔH = -6.24 kcal/mol, ΔS = 7.68 e.u. for site 1 and ΔH = -3.82 kcal/mol, ΔS = 8.27 e.u. for site 2. b) ITC results for BPH-640 binding to DPPS. ΔH = -2.71 kcal/mol, ΔS = 19.29 e.u. c) ΔS/ΔH plot for bisphosphonates binding to FPPS (blue squares) with BPH-629 binding to Rv3378c (red squares, indicated) and BPH-640 binding to DPPS (green square, indicated). Binding of BPH-629 (to Rv3378c) and BPH-640 (to DPPS) are overwhelmingly enthalpy driven (grouped in the blue ellipse), as found with other neutral aromatic side-chain bearing bisphosphonates, binding to FPPS[32].

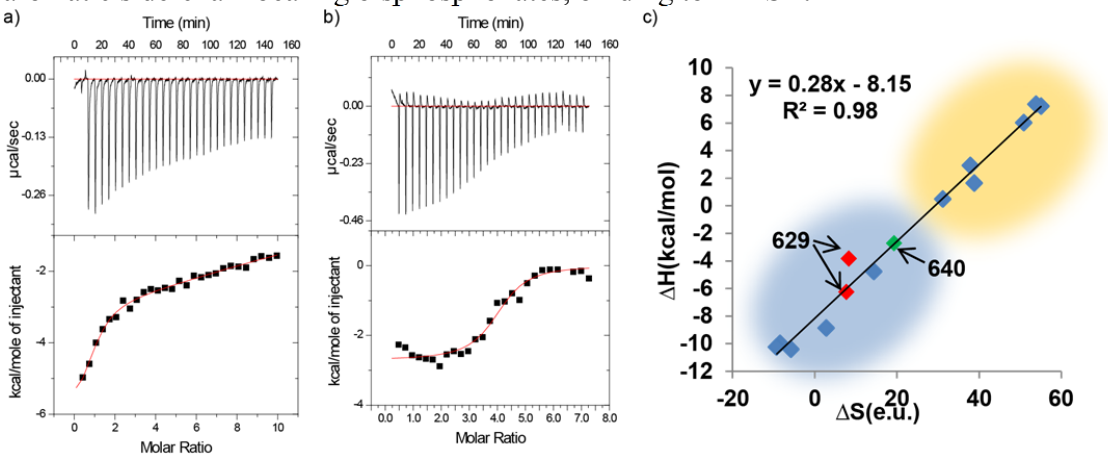


Figure 6.9 Inhibitor bound structures of Rv3378c and DPPS. a) Dimeric Rv3378c structure with two bound BPH-629 molecules (shown with spheres) in each monomer (4 BPH-629 molecules in the dimer). Because there is only one molecule in an asymmetric unit, the dimer is constructed from the observed monomer structure and its symmetry partners. b) Front view of Rv3378c dimer. Monomer A and monomer B are shown as a yellow cartoon and a pink surface, respectively. c) Close-up view of the structure of Rv3378c bound to bisphosphonate inhibitor BPH-629 (thick orange sticks, electron densities contoured at 1.5σ). The BPH-629 molecules both coordinate to the catalytic Mg^{2+} (green sphere), which in turn coordinates to D34 (yellow sticks) from monomer A. BPH-629 (1) is buried in the substrate binding site in monomer A, formed by hydrophobic residues shown as orange lines. The hydrophobic moiety of BPH-629 (2) is stabilized by Y259, S260 and L267 from monomer B (magenta lines). d) Dimeric DPPS structure with one bound BPH-640 molecule (shown as spheres) in each monomer. e) Front view of DPPS dimer. Monomer A and monomer B are shown as a green cartoon and a cyan surface, respectively. f) Close-up view of the structure of DPPS bound to bisphosphonate inhibitor BPH-640 (thick green sticks, electron densities contoured at 1.5σ). The hydrophobic moiety of BPH-640 is stabilized by residues from monomer A (green lines) and monomer B (cyan lines).

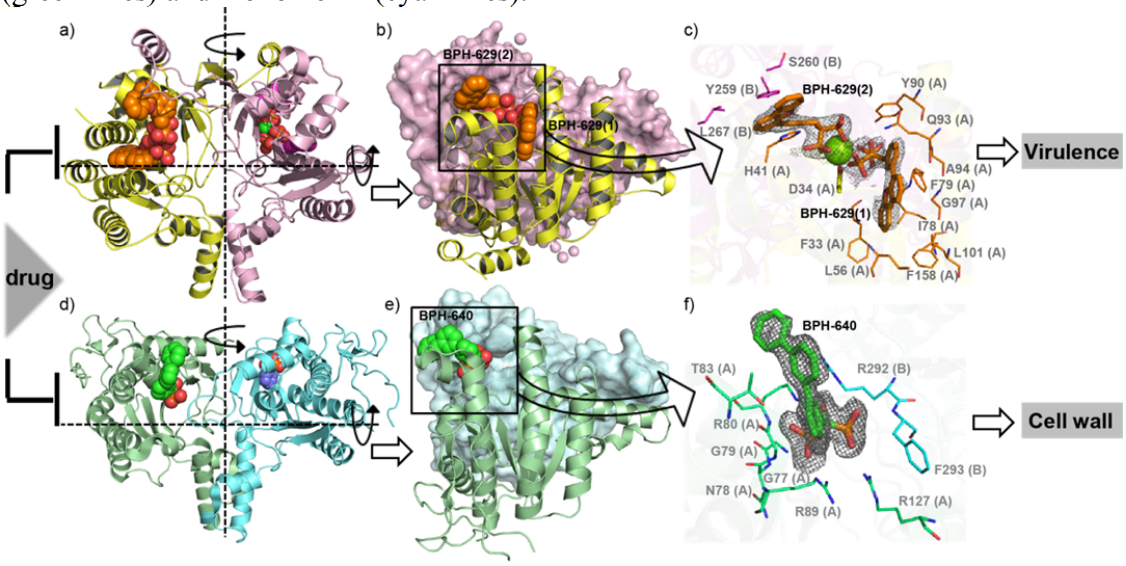


Figure 6.10 Stereo-views of Rv3378c structures shown in Figures 6.9a to 6.9c.

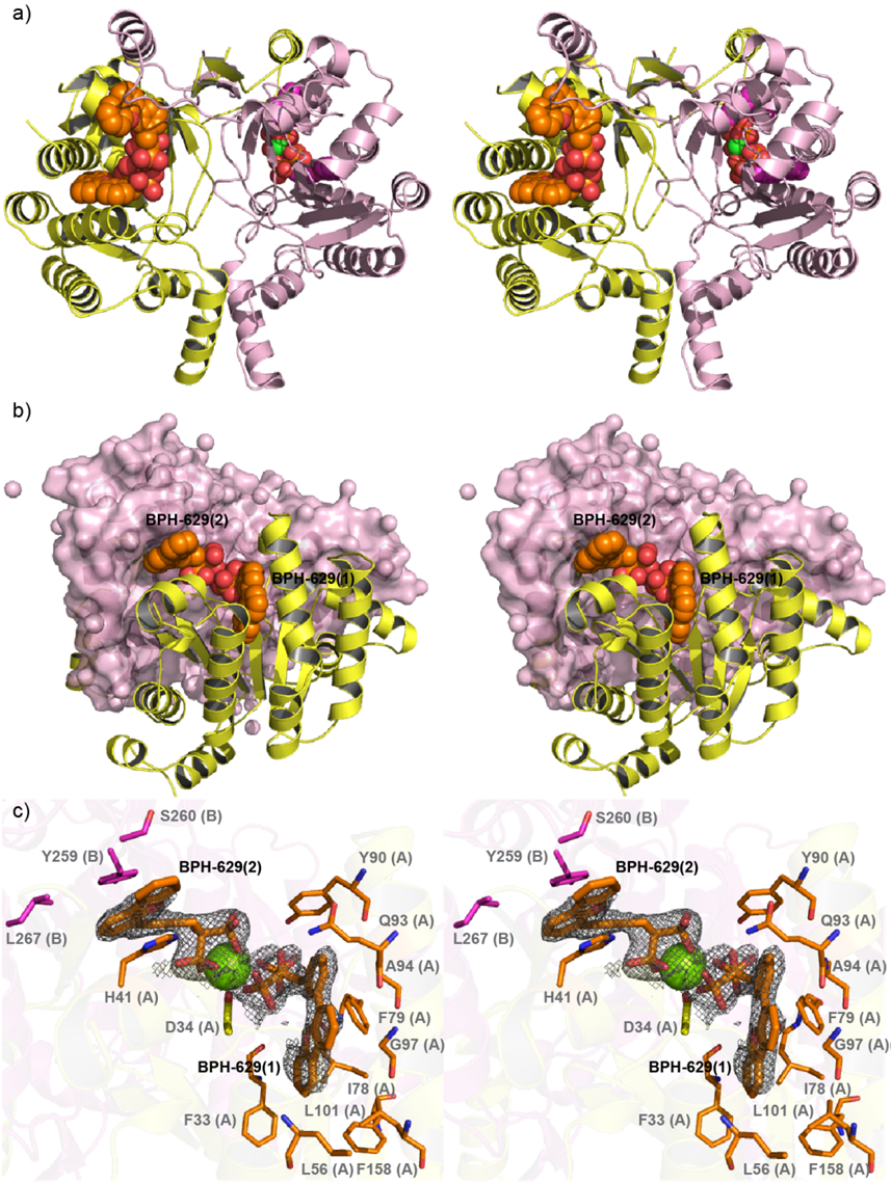
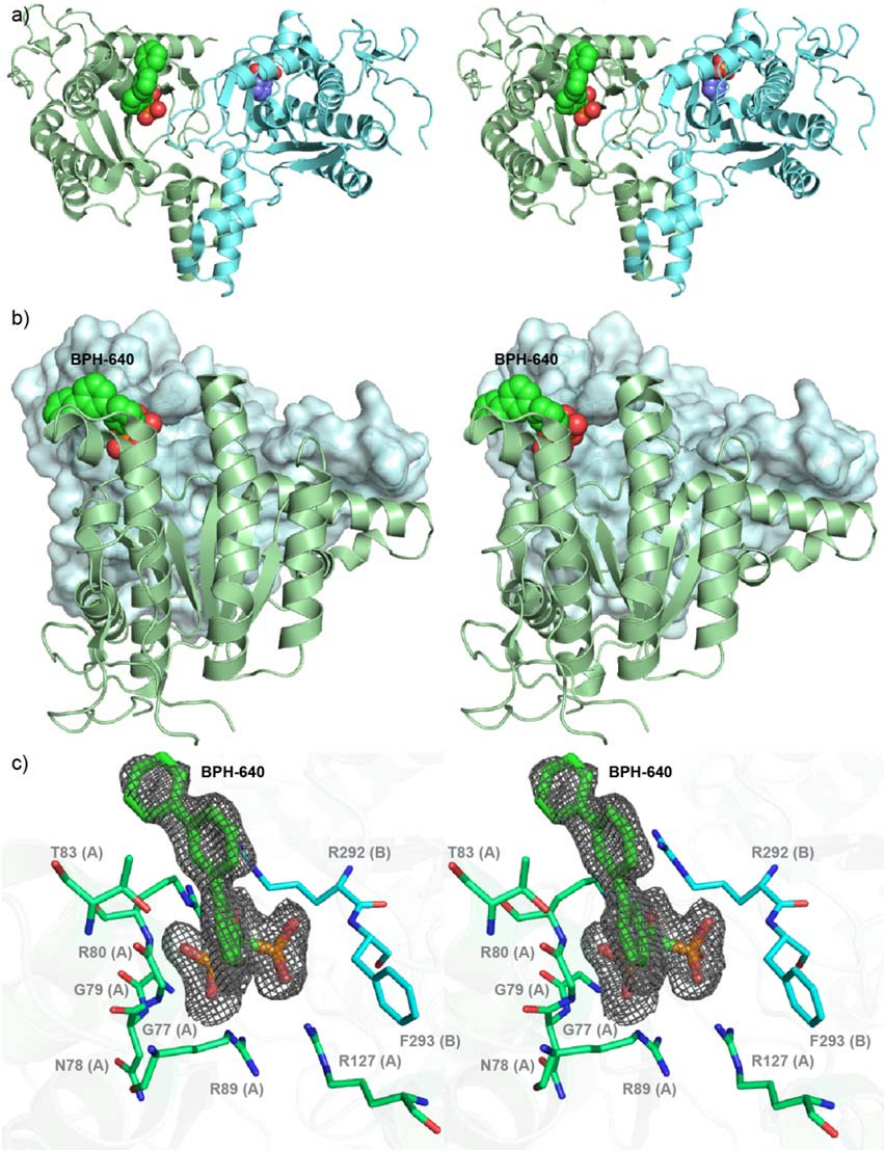


Figure 6.11 Stereo-views of DPPS structures shown in Figures 6.9d to 6.9f.



6.7 References

(1) Ernst, J. D. *Nat. Rev. Immunol.* **2012**, *12*, 581.

(2) Lienhardt, C.; Glaziou, P.; Uplekar, M.; Lonnroth, K.; Getahun, H.; Raviglione, M. *Nat. Rev. Microbiol.* **2012**, *10*, 407.

(3) Zumla, A.; Hafner, R.; Lienhardt, C.; Hoelscher, M.; Nunn, A. *Nat. Rev. Drug. Discov.* **2012**, *11*, 171.

(4) Koul, A.; Arnoult, E.; Lounis, N.; Guillemont, J.; Andries, K. *Nature* **2011**, *469*, 483.

(5) Liu, G. Y.; Essex, A.; Buchanan, J. T.; Datta, V.; Hoffman, H. M.; Bastian, J. F.; Fierer, J.; Nizet, V. *J. Exp. Med.* **2005**, *202*, 209.

(6)Liu, C. I.; Liu, G. Y.; Song, Y.; Yin, F.; Hensler, M. E.; Jeng, W. Y.; Nizet, V.; Wang, A. H.; Oldfield, E. *Science* **2008**, *319*, 1391.

(7) Nakano, C.; Hoshino, T. *Chembiochem* **2009**, *10*, 2060.

(8) Hoshino, T.; Nakano, C.; Ootsuka, T.; Shinohara, Y.; Hara, T. *Org. Biomol. Chem.* **2011**, *9*, 2156.

(9) Nakano, C.; Ootsuka, T.; Takayama, K.; Mitsui, T.; Sato, T.; Hoshino, T. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 75.

(10) Rudi, A.; Aknin, M.; Gaydou, E.; Kashman, Y. *J. Nat. Prod.* **2004**, *67*, 1932.

(11) Maugel, N.; Mann, F. M.; Hillwig, M. L.; Peters, R. J.; Snider, B. B. *Org. Lett.* **2010**, *12*, 2626.

(12)Spangler, J. E.; Carson, C. A.; Sorensen, E. J. *Chem. Sci.* **2010**, *1*, 202.

(13) Pethe, K.; Swenson, D. L.; Alonso, S.; Anderson, J.; Wang, C.; Russell, D. G. *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 13642.

(14) Mann, F. M.; Xu, M.; Chen, X.; Fulton, D. B.; Russell, D. G.; Peters, R. J. *J. Am. Chem. Soc.* **2009**, *131*, 17526.

(15) Mann, F. M.; Prisic, S.; Hu, H.; Xu, M.; Coates, R. M.; Peters, R. J. *J. Biol. Chem.* **2009**, *284*, 23574.

(16) Hendlich, M.; Rippmann, F.; Barnickel, G. *J. Mol. Graph. Modell.* **1997**, *15*, 359.

(17) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177.

(18) Valdar, W. S. *Proteins* **2002**, *48*, 227.

(19) Wang, W.; Dong, C.; McNeil, M.; Kaur, D.; Mahapatra, S.; Crick, D. C.; Naismith, J. H. *J. Mol. Biol.* **2008**, *381*, 129.

(20) Zhang, Y.; Cao, R.; Yin, F.; Hudock, M. P.; Guo, R. T.; Krysiak, K.; Mukherjee, S.; Gao, Y. G.; Robinson, H.; Song, Y.; No, J. H.; Bergan, K.; Leon, A.; Cass, L.; Goddard, A.; Chang, T. K.; Lin, F. Y.; Van Beek, E.; Papapoulos, S.; Wang, A. H.; Kubo, T.; Ochi, M.; Mukkamala, D.; Oldfield, E. *J. Am. Chem. Soc.* **2009**, *131*, 5153.

(21) Morphy, J. R. H., C. J. *Designing Multi-Target Drugs* Royal Society of Chemistry: London, 2012.

(22) Lin, F. Y.; Zhang, Y.; Hensler, M.; Liu, Y. L.; Chow, O. A.; Zhu, W.; Wang, K.; Pang, R.; Thienphrapa, W.; Nizet, V.; Oldfield, E. *ChemMedChem* **2012**, *7*, 561.

(23) Otwinowski, Z.; Minor, W. *Macromolecular Crystallography, Pt A* **1997**, *276*, 307.

(24) Brunger, A. T. *Acta Crystallogr. D Biol. Crystallogr.* **1993**, *49*, 24.

(25) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.;

Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213.

(26) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905.

(27) Emsley, P.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126.

(28) Wang, W.; Dong, C.; McNeil, M.; Kaur, D.; Mahapatra, S.; Crick, D. C.; Naismith, J. H. *J. Mol. Biol.* **2008**, *381*, 129.

(29) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 386.

(30) Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1355.

(31) Nakano, C.; Ootsuka, T.; Takayama, K.; Mitsui, T.; Sato, T.; Hoshino, T. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 75.

(32) Yin, F. L.; Cao, R.; Goddard, A.; Zhang, Y. H.; Oldfield, E. *J. Am. Chem. Soci.* **2006**, *128*, 3524.

Chapter 7: Multi-Target Drug Discovery for Tuberculosis and Other Infectious Diseases

7.1 Notes and Acknowledgements

K. Li synthesized all compounds. L. A. Schurig-Briccio, X. Feng, A. Upadhyay, V. Pujari, B. Lechartier, F.L. Fontes, H. Yang, G. Rao, W. Zhu, A. Gulati, J. H. No, G. Cintra, S. Bogue, Y.-L. Liu, K. Molohon, P. Orlean, D. A. Mitchell, L. Freitas-Junior, F. Ren, H. Sun, T. Jiang, Y. Li carried out biochemistry, enzymatic and cell-based assays. X.Feng, R.-T. Guo, S. T. Cole, R. B. Gennis, D. C. Crick and E. Oldfield analyzed the data. I sincerely thank all the colleagues and collaborators.

This work was supported by the United States Public Health Service (National Institutes of Health grants GM065307, CA158191, HL016101 and AI049151); the National Basic Research Program of China (grants 2011CB710800 and 2011CBA00805); the National Research Foundation of Korea (NRF) funded by the Korean government (MSIP, no. 2007-00559), Gyeonggi-do (no. K204EA000001-09E0100- 00110), and KISTI; and by a grant from the NIH Director's New Innovator Award Program (DP2 OD008463 to D.A.M.); the European Community's Seventh Framework Programme (Grant 260872, S.T.C.), and the Fondation Jacqueline Beytout (B.L.) X.F. was an American Heart Association, Midwest Affiliate, Predoctoral Fellow (grant 13PRE14510056).

This chapter was reproduced in part with permission from K. Li, L. A. Schurig-Briccio, X. Feng, A. Upadhyay, V. Pujari, B. Lechartier, F.L. Fontes, H. Yang, G. Rao, W. Zhu, A. Gulati, J. H. No, G. Cintra, S. Bogue, Y.-L. Liu, K. Molohon, P. Orlean, D. A. Mitchell, L. Freitas-Junior, F. Ren, H. Sun, T. Jiang, Y. Li, R.-T. Guo, S. T. Cole, R. B. Gennis, D. C. Crick, and E. Oldfield. *J. Med. Chem.* February 25, 2014. K. Li, L. A. Schurig-Briccio and X. Feng contributed equally to this work.

Abbreviations used: UPPS, undecaprenyl diphosphate synthase; FPPS, farnesyl diphosphate synthase; GGPPS, geranylgeranyl diphosphate synthase; CrtM, *S. aureus* dehydrosqualene synthase; MK, menaquinone; MenA, 1, 4-dihydroxy-2-naphthoate polyprenyl transferase; MenG, 2-polyprenyl-1,4-naphthoquinone methyltransferase;

TMM, trehalose monomycolate; IC_{50}, half maximal inhibitory concentration; MIC_{90}, minimum *inhibitory* concentration required to *inhibit* the growth of 90% of organisms; PMF, proton motive force; Mt, *M. tuberculosis*; Ms, *M. smegmatis*; Sa, *S. aureus*; Bs, *B.* subtilis; Ec, *E. coli*; Sc, *S. cerevisiae*; Ca, *C. albicans*; Pf, *P. falciparum*; MTT, 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl -2*H*-tetrazolium bromide; IMV, inverted membrane vesicles; TI, therapeutic index; CCCP, *meta*-chlorophenylcarbonylcyanide phenylhydrazone;, ACMA, 9-amino-6-chloro-2-methoxyacridine; TLC, thin layer chromatography; THF, tetrahydrofuran.

7.2 Introduction

Antibiotic resistance is a public health problem that, arguably, has the potential to destroy the efficacy of all antibiotics in the next 10-20 years [1]. There is, therefore, an urgent need for new drugs-especially ones that might be more "resistance-resistant". One possible approach to achieving this goal is to move away from targeting the direct killing of pathogens to inhibiting their virulence, since this might lead to a decreased "life or death" pressure on the organism to develop resistance[2]. A second approach would be to develop more drugs that target pathogen cell membranes. An example of this type of drug would be the anti-fungal amphotericin[3], which functions by binding to ergosterol (which is not present in human cell membranes). A third and well known approach is to employ combination therapies[4], although the problems associated with finding two new drugs active against two new targets are clearly significant. A fourth approach is to use "multi-targeting" or "polypharmacology" in which a single drug has more than one target[5, 6]. This could involve "series inhibition", in which targets would be in the same metabolic pathway (Figure 7.1, left); "parallel inhibition", in which the targets would be unrelated but an inhibitor might mimic a common substrate or would affect e.g. a membrane function (Figure 7.1, middle); or "network inhibition", in which many targets in series and/or in parallel could be involved (Figure 7.1, right). While challenging, many drugs that have been effective in mono-therapy do in fact have multiple targets[5] while single-target drugs (many of which are used in treating tuberculosis) are often only effective in combination therapies.

In this work, we consider the mechanism(s) of action of the new anti-tuberculosis drug SQ109 (**1**) (Chart 7.1), currently in Phase II clinical trials[7]. This drug candidate appeared of interest since it contains a C_{10} isoprenoid (geranyl) side-chain together with a strongly basic (ethylenediamine) fragment-a likely cationic center-suggesting that it might act as a carbocation isostere for a transition state/reactive intermediate in isoprenoid biosynthesis[8] and, as with other inhibitors of isoprenoid biosynthesis, it might be involved in multi-targeting. **1** was developed in a synthesis/screening program[9] in which ~64,000 ethylenediamine analogs of the anti-tuberculosis drug ethambutol (**2**) were synthesized. **1** was ~ 4x more active than any of the other leads developed, having a minimum inhibitory concentration (MIC) of ~0.7-1.56 μM against *M. tuberculosis* (H37Rv, Erdman and drug-resistant strains) and insights into its mode of action recently became available when the target of SQ109 was proposed[10] to be MmpL3, a trehalose monomycolate (TMM) transporter, an essential membrane protein that transports TMM into the cell envelope. This conclusion was based on the observation that several *M. tuberculosis* mutants produced via serial passage with several **1**-like inhibitors had mutations in the *mmpL3* gene and cross-resistance to **1**, although these latter effects were rather small[10]. More intriguingly, no spontaneous resistant mutants were obtained when using **1**, suggesting the possibility of multiple targets[10-11]. That idea is supported by the observation that **1** also has activity against other bacteria, e.g. *Helicobacter pylori*[12] as well as against the yeast *Candida albicans*[13], neither of which possess the *mmpL3* gene, suggesting again that other **1** targets are harbored by these organisms, and potentially, by *M. tuberculosis*.

In this work, we synthesized a series of analogs of **1** in which we varied the adamantane head-group and the ethylenediamine linker (varying the possible charge centers), Chart 7.1. All compounds were screened against a panel of bacteria (*M. tuberculosis*, *M. smegmatis*, *Escherichia coli*, *Staphylococcus aureus* and *Bacillus subtilis*), two yeasts (*Saccharomyces cerevisiae* and *Candida albicans*), a malaria parasite (*Plasmodium falciparum*) and a human cell line (MCF-7), to establish anti-bacterial, anti-fungal and anti-malarial structure-activity relationships and to assess mammalian cell toxicity. In addition, we investigated a subset of compounds for activity against a series

of putative targets, isoprenoid biosynthesis enzymes, while also investigating the effects of these compounds on respiration, ATP synthesis, and the proton motive force (PMF).

7.3 Results and Discussions

Only one cationic center is needed for potent activity in mycobacteria. To investigate which features contribute to the activity of **1** against a broad range of organisms we synthesized **1** and the 11 analogs (**3-13**) shown in Chart 7.1 in which the ethylenediamine linker was replaced by ethanolamine, choline, propanolamine, ethylene glycol or glycolic amide moieties (providing linkers with potentially, 0, 1 or 2 positive charges), as well several "head-group" analogs in which the adamantyl group was varied. The geranyl "side-chain" was kept constant. Full synthesis and characterization details are given in the Supporting Information. As expected, **1** had potent activity against *M. tuberculosis* with an MIC_{90} of ~ 0.1-0.2 μg/mL (Table 7.1). Interestingly, the *N*-geranyl ethanolamines **3** and **13** were more potent (MIC_{90} values as low as 0.02-0.05 μg/mL), indicating that the presence of two nitrogens was not essential for activity. The *O*-geranyl ethanolamine derivative (**4**) had similar activity to **1** (~0.2 μg/mL), Table 7.1. With the *N*-geranyl ethanolamines (**3, 13**), activity was about 30-fold higher than with ethambutol (**2**). The ethylene glycol (**5**) was far less active (50 μg/mL), as was the glycolic amide **9** (12 μg/mL). Thus, the most active compounds all contain one strongly basic nitrogen in the linker region, with most activity being found with the two *N*-geranyl ethanolamines **3** and **13** (Table 7.1), with **3** being ~4-5 times more active than was **1**.

We then tested all 12 compounds against *M. smegmatis*. The results (Table 7.1) show that overall activity against *M. smegmatis* is less than that observed against *M. tuberculosis*, as can also be seen in the "heat-map" shown in Figure 7.2A. There is, however, a very high correlation coefficient (R^2 = 0.9, Figure 7.2B) between the pIC_{50} (=$-\log_{10}IC_{50}$) values for *M. tuberculosis* and *M. smegmatis*, indicating a similar mechanism of action, leading to our use of *M. smegmatis* in several mechanism of action studies, described below.

The cationic inhibitors exhibit broad anti-bacterial, anti-fungal, and anti-malarial activity. All 12 compounds (**1,3-13**) were then tested against three other

bacteria: *S. aureus*, *B. subtilis* and *E. coli*, Table 7.1. With *S. aureus* (the methicillin resistant *S. aureus* (MRSA) USA300 strain), **1** itself had little activity, however, the *N*-geranyl ethanolamines **3, 12, 13** and the N-geranyl propanolamine **6** all had MIC_{90} values of ~8 μg/mL, while the other analogs were much less active (MIC_{90} > 32 μg/mL). A similar pattern of activity was seen against *B. subtilis*, with the three ethanolamines (**6, 12, 13**) exhibiting the highest levels of activity. In addition, unlike with *S. aureus*, **1** itself had activity (7.6 μg/mL, Table 7.1). With *E. coli*, **1**, the *N*-geranyl ethanolamine (**3**) and the *N*-geranyl propanolamine (**6**) were all quite active with IC_{50} values of ~2-3 μg/mL, Table 7.1. Moreover, there was a modest correlation (R^2 = 0.5) between the *M. tuberculosis* (or *M. smegmatis*) $pMIC_{90}$ values and those found with *E. coli*, Figure 7.2B. These results again indicate that at least one basic amine-most likely a cationic center -is required for best activity, plus, there must be a target or targets other than MmpL3 in *E. coli* since the *mmpL3* gene is absent in this organism. Bioinformatics searches did locate uncharacterized *mmpL3*-like genes in *S. aureus* and *B. subtilis,* but it remains to be seen if the corresponding proteins are targeted by our compounds.

We next tested all 12 compounds (**1,3-13**) for activity against *S. cerevisiae* and *C. albicans*. As can be seen in Table 7.1 and Figure 7.2A, **1** and the ethanolamines **3, 6, 11** and **13** had activity in the 1-3 μg/mL range with **1** and **13** being the most potent, having an IC_{50} ~ 1 μg/mL. With *C. albicans*, **1** was most active, followed by the ethanolamines **3** and **13**. Not unexpectedly, there was a high correlation between the pIC_{50} values seen between *S. cerevisiae* and *C. albicans* (R^2 = 0.8; Table 7.1 and Figure 7.2B). A modest correlation between the pIC_{50} values for the yeasts and *E. coli* or *B. subtilis* is also apparent (R = 0.6-0.7, Figure 7.2B), suggesting the possibility of target conservation between fungi and bacteria. Since a recognizable *mmpL3* gene is absent in the fungi, these results again indicate an alternate target or targets. These results also lead to the idea that there could be additional targets in *M. tuberculosis*, which would help explain the very low MIC values observed and the inability to induce resistance via serial passage, as noted by Tahlan et al[10]., although multiple-targeting does not necessarily guarantee improved potency. The results with the bacteria and fungi then suggested the possibility that the growth of other organisms (protozoa) might also be inhibited by **1**, or its analogs.

To evaluate anti-protozoal activity we screened all 12 compounds (**1,3-13**) against the intra-erythrocytic form of the malaria parasite, *Plasmodium falciparum*. As can be seen in Table 7.1, **1** had a ~1 μg/mL activity against *P. falciparum*, and the three ethanolamines (**3, 4** and **6**) also had good activity. As viewed on the heat-map (Figure 7.2A), inhibition of *M. tuberculosis* cell growth is strongest, but is followed by *P. falciparum* (in the intra-erythrocytic assay) and in each case where there is activity against *P. falciparum* (**2, 3, 4, 6, 7**), the inhibitors (Chart 7.1) are expected to carry a +1 charge, as with the best *M. tuberculosis* growth inhibitors. When compared to growth inhibition results with a human cell line (MCF-7; Table 7.1), we see that activity against the human cells is much weaker than against *P. falciparum* and, of course, against *M. tuberculosis*. We calculate a therapeutic index (TI), defined as:

$$\mathrm{TI} = \frac{\mathrm{IC_{50}\ (human\ cell\ line)}}{\mathrm{IC_{50}\ (pathogen)}}$$

of ~18 for **1** against *P. falciparum* and ~40 for **3**, while for *M. tuberculosis* we find TI = 120 (**1**) and TI = 900 (**3**), suggesting that these and related analogs may also be promising *P. falciparum* drug leads. We next sought to explore what the additional targets for these compounds might be in cells that lack MmpL3.

Possible protein targets for SQ109 and its analogs. The general patterns of activity seen with the compounds described above have some similarities across the diverse organisms investigated in that at least one cationic center, or perhaps more importantly a protonatable nitrogen, is required for activity. In *M. tuberculosis*, **1** is thought to act by inhibiting MmpL3, a TMM transporter [10], although as noted by Tahlan *et al.*, other targets could also be involved. This seems quite likely since in most cases these other organisms lack *mmpL3* or a clearly identifiable ortholog, and do not utilize TMM, as is also the case with *H. pylori*[12]. Given that protonated geranylamines might be good isosteres for transition states or reactive intermediates in enzymes involved in isoprenoid biosynthesis (Scheme 7.1), we investigated if **1** could inhibit any of the following enzymes: *M. tuberculosis cis*-farnesyl diphosphate synthase (Rv1086);

M. tuberculosis cis-decaprenyl diphosphate synthase (Rv2361); *P. vivax* geranylgeranyl diphosphate synthase (GGPPS); *S. aureus* and *E. coli* undecaprenyl diphosphate synthases (UPPSs), *S. aureus* farnesyl diphosphate synthase (FPPS) and

human GGPPS. In essentially all cases, IC_{50} values were ≥ 50μM. The exception was human GGPPS which was inhibited by **1** (the only compound with two basic groups) with a 4.5 μM IC_{50}. These enzymes are all so-called *cis* or *trans*- "head-to-tail" prenyltransferases[14] and the presence of the two (as opposed to one) hydrophobic domains (in addition to the cationic center) might not be required for enzyme inhibition. There are, however, other prenyl transferases that might be targeted in which two hydrophobic domains - together with a carbocation center - would better mimic transition states/reactive intermediates. These would include the so-called "head-to-head" prenyl transferases, as well as some of the enzymes involved in quinone biosynthesis. There are demonstrated or putative head-to-head prenyl transferases in *M. tuberculosis* (Rv3397c), *M. smegmatis, (*Mycsm_04912), *S. aureus* (CrtM), *B. subtilis* (YisP), *S. cerevisiae* (squalene synthase, SQS) and *C. albicans* (SQS) and in humans (SQS), but no homologous proteins can be found by standard BLAST searches in *P. falciparum*. The products (where known) of these enzymes vary and not all are essential for survival *in vitro*. Nevertheless, we tested a subset of compounds (**1, 3**, **4**) for activity against either SaCrtM or human SQS finding only weak activity (~100 μM) in all cases. These results support the notion that the head-to-head prenyl transferases are unlikely cell growth inhibition targets of our compounds in these organisms.

The other obvious candidates are the enzymes involved in quinone biosynthesis (Scheme 7.2) or quinone utilization. We thus next investigated the two quinone biosynthesis enzymes, MenA and MenG, both of which are likely to utilize cationic transition states/reactive intermediates during catalysis. MenA (EC 2.5.1.74, 1, 4-dihydroxy-2-naphthoate polyprenyl transferase) catalyzes the isoprenylation of 1, 4-dihydroxy-2-naphthoic acid by long chain isoprenoid diphosphates[15], Scheme 7.2, and MenA is of interest as an *M. tuberculosis* drug target[16, 17, 18]. In an initial set of experiments we tested three potent *M. tuberculosis* and *M. smegmatis* growth inhibitors (**1, 3** and **13**) in the *M. smegmatis* MenA (MsMenA) membrane fragment inhibition assay described previously[16, 17, 18] finding IC_{50} values of ~6 μM (Table 7.2). Typical dose response curves are shown in Figure 7.3A. While this assay revealed only modest activity, the observation that MenA activity was in fact inhibited by the three most potent

inhibitors is a potentially important one since this inhibition might be expected to inhibit electron transfer/ATP synthesis, of particular importance in non-replicating/persister cells[18] and to contribute to cell growth inhibition beyond that seen with MmpL3 inhibition alone. What was also of interest was that **1** had similar activity (IC_{50}=9 M, in the same assay as used here[18]) against MenA to that we reported previously with Ro 48-8071, a lipophilic amine that decreases menaquinone biosynthesis and blocks *M. tuberculosis* as well as *M. smegmatis* cell growth. These growth inhibition effects with Ro 48-8071 (as well as the inhibition of respiration) were reported previously to be reversed in both organisms by addition of 400 μM menaquinone-4 (MK-4) or phylloquinone to the medium[19].

A second possible target is MenG (EC 2.1.1.163, 2-polyprenyl-1,4-naphthoquinone methyltransferase) which carries out the *S*-adenosylmethionine (SAM)-dependent methylation of demethylmenaquinone (the product of the MenA reaction). As with MenA, the MenG reaction is inhibited by **1** and the potent ethanolamine analogs **3**, **13** (Table 7.2 and Figure 7..3B and C), with IC_{50} in the 6-13 μM range. Unlike the *C*-alkylations with prenyl diphosphates, the MenG reaction uses SAM (as a C_1 source), and Mg^{2+} is not required. With **1** binding to MenG, the cationic center in the inhibitor might mimic a cationic transition state/reactive intermediate, although another possibility is that the cationic center simply mimics the SAM *S*-methyl sulfonium group. Thus, both MenA and MenG are inhibited *in vitro* by **1** and its analogs, which can be expected to supplement MmpL3-based inhibition in the mycobacteria, as well as provide alternative targets in some of the organisms that lack the *mmpL3* gene. Moreover, inhibition of two sequential targets (series inhibition) in a biosynthetic pathway can often be quite effective since the product of the first reaction is the substrate for the second reaction[20].

We next used an expressed *E. coli* MenA (hereafter EcMenA) detergent-based assay to obtain inhibition data for all 12 inhibitors, Table 7.2 and Figure 7.4. Interestingly, the most potent inhibitor was **3** (IC_{50}= 400 nM), and **3** was also the most potent inhibitor of *M. tuberculosis* cell growth (and, within experimental error, of *E. coli* cell growth, Table 7.1). We additionally found that there was a moderate correlation between *E. coli* cell growth inhibition and EcMenA inhibition with an R^2 = 0.43 (using $pIC_{50} = -\log_{10} IC_{50}$, both values in μM) values, suggesting that MenA inhibition may be involved in cell

growth inhibition. As described below, the experimental *vs.* predicted *E. coli* cell growth inhibition correlation increased to R^2 = 0.77 with the incorporation of a second experimental parameter, pH collapse.

The structure of MenA is not known but it is predicted to be a trans-membrane protein containing ~9 α-helices, as shown in Figure 7.5A[21]. Using modern structure prediction programs such as Phyre2[22] that are secondary-structure based, MenA is predicted (Figure 7.5B) to adopt basically the same all α-helical fold as found in farnesyl diphosphate synthase and CrtM (the *S. aureus* dehydrosqualene synthase), but where 1 N and 2 C-terminal helices (trans-membrane helices 1, 8 and 9 in Figure 7.5A) are not modeled, Figure 7.5B. 198 residues (68%) are, however, modeled at a predicted >90% accuracy, and the predicted structure has closest similarity to the crystal structure of farnesyl diphosphate synthase from *Methylococcus capsulatus* (PDB ID code # 3TS7), although remarkably, there is only a 10% residue identity. The 1st and 2nd aspartate-rich motifs essential for Mg^{2+} binding and catalysis in FPPS and CrtM are located in very similar regions in the MtMenA model, as shown in the superposition with CrtM in Figure 7.5C (orange spheres =conserved Asps in EcMenA model; blue spheres= Asp-rich motif in CrtM). This then suggests, based on the **1**-CrtM X-ray structure[23], the binding sites for **1** (pink) shown in Figure 7.5C. The two Asp-rich domains in MenA are also highly conserved, as shown by a SCORECONS[24] analysis (Table 7.3). Although only a computational prediction, it is of interest that the highest scoring Phyre2 prediction is found with a prenyl transferase enzyme that is known to utilize a carbocation mechanism, consistent with the experimental observation that only cationic species inhibit MenA.

Menaquinone rescue experiments. We next measured the activity of **1** against both actively growing (*M. tuberculosis* H37Rv) and non-replicating (streptomycin-starved *M. tuberculosis* 18b)[25] mycobacteria, using a resazurin microplate reduction assay (REMA; Figure 7.6). **1** had a MIC of 0.15 μg/mL against actively replicating H37Rv in this assay, as expected. It also displayed activity against the non-replicating streptomycin-starved 18b strain (where MmpL3/TMM transport is presumably not involved since there is no cell growth), and the effects of **1** on both strains were affected by MK-4 addition (Figure 7.6). In the H37Rv aerobic assay, the MIC shifted from 0.15 μg/mL in the absence of MK-4 (Figure 7.6A) to ~1 μg/mL when the medium was supplemented with 1 mM

menaquinone, consistent with a role for **1** in inhibiting quinone biosynthesis and/or electron transport. As noted above, a remarkably similar effect was seen previously with Ro 48-8071, another lipophilic amine, at 400 μM MK-4[19], for both *M. tuberculosis* and *M. smegmatis*. The effect of **1** against non-replicating (streptomycin-starved 18b) bacteria, as seen by REMA as a decrease in fluorescence (lack of resazurin reduction to the highly fluorescent red resorufin) above a **1** concentration of 1 μg/mL was also blocked by MK-4 addition (Figure 7.6B). The activity of **1** against non-replicating (streptomycin-starved 18b) cells was confirmed by plating and counting CFU after 7 days of drug exposure (Figure 7.6C) with normal 7H9 medium or with 7H9 medium containing 1 mM MK-4. As can be seen in Figure 7.6C, **1** at 1 μg/mL had essentially no effect on (non-replicating) bacterial activity in the presence of MK-4, and only a small effect in the absence of MK-4. However, at 10 μg/mL **1**, while there was again a small effect on activity in the presence of MK-4, cell activity in the absence of MK-4 was reduced by ~4 log units, consistent with a role for **1** in blocking respiration and hence, ATP synthesis.

Respiration, TMM and the PMF. The results described above show that **1** has activity against not only the two mycobacteria (*M. tuberculosis* and *M. smegmatis*), but also against a range of other bacteria, fungi and a protozoan, each of which lack a bioinformatically- recognizable *mmpL3* gene. In *M. tuberculosis* and *M. smegmatis*, inhibition of MenA/MenG would inhibit respiration, resulting in a decrease in ATP biosynthesis. This could help explain how **1** increases the level of TMM given that MmpL3 is a TMM transporter of the RND family of efflux pumps, many of which are powered by the PMF. Restated, **1** might exhibit an indirect action upon the TMM transporter by removing its "power source" (the proton motive force), in addition to directly binding to, and inhibiting, the transporter. This indirect action could be accomplished in one of two ways: (1) blocking respiration (by depletion of menaquinone by inhibition of MenA, MenG, or by directly inhibiting a component of the electron transport chain); or (2) a direct effect on the PMF ($\Delta\psi + \Delta pH$, where $\Delta\psi$ is the membrane potential and ΔpH, the pH gradient). The possibility of the involvement of the PMF is suggested from the results of a number of studies in which lipophilic bases (e.g. amiodarone, local anesthetics and NSAIDS[26, 27, 28]) can act as uncouplers. In addition, there could also be multi-drug targeting affecting MmpL3, MenA, MenG and the PMF,

which would be expected to produce potent inhibition of cell growth/respiration/ATP synthesis, as well as a low rate of resistance.

Respiration and electron transport. In earlier work[17, 19], we showed that several MenA inhibitors, analogs of Ro 48-8071, blocked respiration in *M. tuberculosis* and *M. smegmatis* (as evidenced by inhibiting the reduction of methylene blue); that there was a correlation between cell growth inhibition and respiration inhibition[19], and that the effects of the inhibitors could (at least in part) be reversed by adding MK-4 at the 400 M level. We thus next tested all compounds for their effects on methylene blue reduction, in *M. smegmatis* (Figure 7.7), finding that there was a moderate correlation between $pMIC_{90}$ (=-$\log_{10}MIC_{90}$, MIC_{90} in μM) for cell growth inhibition and the pIC_{50} (=-$\log_{10}IC_{50}$, IC_{50} in μM) for inhibition of whole cell respiration inhibition (R^2=0.55) for all 12 compounds (Tables 7.1 and 7.2).

These results suggest the possibility of a direct effect on electron transport (since the effects observed are rapid-10s of minutes), blocking respiration, consistent with the MK-4 rescue experiments. The nature of the target or targets involved are beyond the scope of this current study, but we did carry out preliminary experiments with **1** against a series of dehydrogenases by monitoring the reduction of the artificial electron acceptor MTT (3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide). We used an *M. smegmatis* membrane preparation and a variety of substrates including NADH (measuring both Complex I and alternative NADH dehydrogenases, NDH-2); deamino-NADH (measuring Complex I activity but not NDH-2); succinate (measuring succinate dehydrogenase); malate (measuring quinone-dependent malate dehydrogenase); and lactate (measuring lactate dehydrogenase). IC_{50} values for **1** were in general ~30 μg/mL, the exception being malate dehydrogenase (IC_{50} = 10 μg/mL), results that suggest that more than one electron-transfer protein may be involved, in cells, with the inhibitors mimicking quinone substrates.

Uncoupler effects in membrane vesicles and in cells. The results presented above show that **1** and its analogs inhibit MenA, MenG, electron transfer proteins and respiration and that MK-4 can rescue cell growth or activity, the relatively high IC_{50}s for enzyme inhibition/respiration (when compared to cell growth inhibition) being suggestive

of multiple targeting. What is of particular interest about all of the results described above is that they seem to point in one direction: respiration, offering a possible explanation for the previous observation that TMM accumulates (with **1**), in *M. tuberculosis*, due to MmpL3 inhibition. This MmpL3 inhibition could be due to direct binding, or a more indirect effect on the PMF/ATP synthesis.

To test the hypothesis that **1** and its analogs might collapse the PMF, we first used *E. coli* inverted membrane vesicles (IMVs), essentially as described by Haagsma[29]. The results obtained with SF6847, one of the most potent uncouplers known, are shown in Figure 7.10A and indicate a very rapid (seconds) collapse in ΔpH, as reported by Haagsma et al[29]. The same effect was seen with **1** and analogs that have potent activity in cell growth inhibition, while inactive (diether/amide) analogs (e.g. **5**) had no effect, Figures 7.10B, S4 and S5. Similar results were obtained with both *E. coli* and *M. smegmatis* vesicles. The effects on the collapse in ΔpH were seen in vesicles in which the pH gradient was driven by either ATP hydrolysis, or by electron transport in the presence of succinate or NADH. Using Oxonol VI as a probe, we also found that Δψ in *E. coli* IMVs (positive inside) was collapsed (Figure 7.11A) by the same compounds, and there was a correlation between the collapse of the membrane potential and the collapse in ΔpH (using ACMA fluorescence; Table 7.2; R^2=0.79, Figure 7.11B). This is consistent with these lipophilic cations acting as protonophores, carrying protons across the membrane lipid bilayer, with only compounds containing a basic nitrogen supporting the uncoupling activity.

We found similar PMF effects in intact *M. smegmatis* cells in which there was a collapse in Δψ (positive outside), as measured by using DisC3(5) fluorescence, Figure 7.12. Addition of **1** or the potent analogs to *M. smegmatis* cells resulted in an immediate increase in DisC3(5) fluorescence, indicating a collapse of Δψ. As can be seen in Figure 7.12B, **1** collapses the membrane potential in a dose-dependent manner with an "EC_{50}" of ~20 μg/mL. The EC_{50} for one of the most potent cell growth inhibitors **13** was ~15 μg/mL.

In addition to these investigations of Δψ collapse in intact cells, we investigated pH collapse in intact *M. smegmatis* cells, using ^{31}P NMR spectroscopy. Phosphorus NMR chemical shifts are sensitive indicators of local pH values[30]. As can be seen in Figure

7.13A the ^{31}P NMR chemical shift of phosphate inside *M. smegmatis* is ~ 0.35 ppm downfield from external P_i and from these chemical shifts, the internal and external pH values can be determined: results are shown in Figure 7.13B. There is a ΔpH=0.26 (inside more basic) in wild type *M. smegmatis* cells but this pH gradient is collapsed by the uncoupler CCCP (*meta*-chlorophenylcarbonylcyanide phenylhydrazone), the antiporter nigericin, as well as by **1** and **13**, while as expected, the K^+ ionophore valinomycin has no effect. The effects of **1** and the other potent analog thus leads to collapse of Δψ as well as ΔpH in both inverted vesicles and in whole cells. This collapse in the PMF, in cells, can be expected to result in an inhibition of ATP synthesis, as is indeed found experimentally (Figure 7.10D) with the potent lead, **3**. In addition, the collapse of the proton motive force is expected to inhibit activity of the MmpL3/TMM transporter.

Quantitative models for cell growth inhibition. Any quantitative analysis of cell growth inhibition based on enzyme inhibition or another property (e.g. ΔpH collapse) is of course challenging, but it should be possible to use the multi-descriptor approach described previously[31] with, in this case, no purely mathematical descriptors being required. We thus use Equation 1:

$$pIC_{50}\ (cell) = a \bullet pIC_{50}\ (A) + b \bullet pIC_{50}\ (B) + c$$

Equation 1

where $pIC_{50}(A)$ is $-\log_{10}$ (IC_{50}, A) for enzyme or property A (such as MenA inhibition), and B is a second property, e.g. ΔpH collapse. We show by way of examples in Figures 7.14A,B, three-dimensional plots for *E. coli* cell growth inhibition: $\log IC_{50} = f$ (MenA, ΔpH), and for *M. smegmatis*: $\log MIC_{90} = f$ (respiration, Δψ), where we find correlation coefficients of $R^2 = 0.82$ (*E. coli*) and $R^2 = 0.72$ (*M. smegmatis*) values for the experimental-versus-predicted cell growth inhibition results. In previous work we investigated correlations between cell and enzyme activity (pIC_{50}) assays in 10 diverse systems finding[31] on average an $R^2 = 0.32$ for the 10 cell/enzyme correlations, similar to the $R^2 = 0.43$ we find for the MenA alone/*E. coli* cell growth inhibition correlation. Incorporation of the second parameter (the percentage of ΔpH collapse) increases the R^2 to 0.77, suggesting the importance of multi- targeting, in *E. coli*. The correlation is worse

for *M. smegmatis* due, perhaps, to the omission of an MmpL3 term, expected be particularly important in the mycobacteria.

Are there other MenA-like targets? The results presented above show that *E. coli* MenA is inhibited by **1** and its analogs and that there is a correlation between MenA inhibition and cell growth inhibition (in *M. smegmatis*) suggesting that diverse MenAs may be inhibited by these compounds. However, this result is perhaps surprising in that in *E. coli*, MenA is not an essential gene for aerobic growth because UbiA can be used in aerobic respiration. One possibility is that EcUbiA might also be inhibited by **1** (and its analogs). While we have not yet investigated this experimentally, what we have found is that MtMenA, EcMenA, as well as EcUbiA are all predicted (using the Phyre2 program) to have an FPPS/GGPPS-like structure.

In all three cases the structures are predicted to contain a central FPPS/GPPS-like catalytic domain (comprising ~2/3 of the overall amino-acid sequence) that has close similarity to the same two (soluble) prenyl synthases: *Methylococcus capulatus* FPPS and *Lactobacillus brevis* GGPPS. Predicted sequence identity investigations have shown that MenA and UbiA have moderate homology[32], but correlations with FPPS and GGPPS were not made in those studies because the actual sequence identities are very low, about 10-16%. However, secondary-structure based algorithms do permit accurate structure predictions, even when residue identities are low.

These previous bioinformatics studies also demonstrated that another class of proteins, protoporphyrin IX farnesyl transferases (e.g. heme o synthase) have significant sequence homology to the MenA/UbiA proteins – and all 3 classes of proteins are Mg^{2+} dependent prenyl transferases. Once again, the structure of heme o synthase is not known, but is predicted to be another 9-helix trans-membrane protein with a central FPPS/GGPPS-like core suggesting that MenA, UbiA, as well as protoporphyrin IX farnesyl transferases are all likely to be inhibited by **1** and related systems.

MmpL3 and MmpL11 as targets. MmpL3 is thought to be a target for **1** (and diverse other inhibitors[10-11, 33]), blocking TMM transport. It has also been shown that MmpL3 together with a related protein, MmpL11, are associated with heme uptake[34].

The X-ray structures of MmpL3 and MmpL11 have not been reported. However, both are membrane proteins and are predicted to have 11-12 *trans*-membrane helices[21]. Using the Phyre2 program[22] we find with MmpL3 that 653 residues (69% coverage) are predicted with 100% confidence to have the structure shown in Figure 7.15A and with MmpL11, 642 residues (66% coverage) are predicted with 100% confidence to have the structure shown in Figure 7.15B. Both structures are very similar to those found in cation efflux pumps such as CusA (PDB ID code 3ko7) and multi-drug efflux pumps such as the acriflavin resistance protein B (AcrB; PDB ID code 1oy8), although the C-terminus (~1/3 of the total protein) is not modeled in either MmpL3 or MmpL11. The transmembrane hydrophobic domains are shown in Figures 7.15C, D (in white/light orange). **1** as well as several other inhibitors[10-11] (Figure 7.15E) have been proposed to target MmpL3 (detected by sequencing mutants that arose under drug pressure), but the sites of these mutations, shown as blue spheres in Figure 7.15F, are spread throughout the protein, suggesting perhaps, multi-site targeting of MmpL3/11 as an additional basis for the lack of resistant mutations with **1**.

A multi-target model for anti-infective activity. We show in Figure 7.16 a summary of the proposed sites of action for **1** and its analogs in *M. tuberculosis* and in *M. smegmatis*. Some of these targets are also present in the other pathogens investigated, but not in human cells. In addition to its previously proposed role in targeting MmpL3, **1** and its analogs also inhibit MenA and MenG and, as described above, the inhibition of *M. tuberculosis* cell growth or activity is rescued by MK-4. We also find that the PMF is inhibited by the most active compounds, which act as protonophores/uncouplers. This results in a decrease in ATP synthesis and, we propose, decreased activity of MmpL3/11, helping explain the accumulation[10] of TMM (with **1**).

This multiple-targeting is perhaps best thought of as involving network inhibition in which both series and parallel paths are involved (Figure 7.1C) since at least in the mycobacteria, MenA, MenG, electron transport, ΔpH, $\Delta\psi$ and MmpL3 (and presumably other pumps dependent on the PMF) can all be affected. There are, of course, likely to be differences in the mechanisms of action of different inhibitors in different organisms (and in the same organisms under different growth conditions), although effects on the PMF

are expected to be quite common since they are based on more "physical" properties, rather than purely enzyme inhibition. The uncoupling effects we observe could also help explain the growth inhibition seen in human cell lines, as could inhibition of the human MenA, UbiA, and MenG/UbiE orthologs: UbiAD1, CoQ2 and CoQ3.

Also of interest are the likely differences in timescales (and concentrations) for the different reactions involved. The effects on the collapse in $\Delta\psi$ and ΔpH are very rapid-on the seconds to minutes timescale and are observed (in vesicle experiments) at low μM concentrations, for the most active species. The effects on respiration as determined by methylene blue reduction (in intact cells) are also rapid, typically observable in minutes, and may reflect the time required for inhibitors to enter the cell and accumulate (since they could also be being actively pumped out). Little is known about the rate of menaquinone turnover, but it is likely that several cell divisions are required for a large reduction in menaquinone levels, so while MenA/MenG inhibition may be rapid, the effects on cell growth may take many hours or (with *M. tuberculosis*) days to occur likewise, since MmpL3 is thought to be involved in cell wall biosynthesis, its inhibition would also be expected to result in observable effects on growth inhibition on a time scale of hours to days.

7.4 Conclusions

The results we have described above are of interest for drug discovery against tuberculosis, as well as against other bacterial, fungal and protozoan pathogens, for several reasons. We synthesized a series of analogs of the anti-tuberculosis drug **1** in which we varied the nature of the ethylenediamine linker to provide cationic, protonatable as well as neutral species, and in addition we varied the adamantyl headgroup. The most active compound against *M. tuberculosis* was ~5x more potent than was **1**, and was also less toxic to an MCF-7 human cell line. We tested all compounds against a panel of bacteria, fungi and a protozoan parasite and the results obtained showed that at least one cationic (or basic) group was essential for activity. The most potent activity was against *M. tuberculosis* (MIC=0.02-0.05 μg/mL) and the intra-erythrocytic form of the malaria parasite, *P. falciparum* (IC_{50} = 30 ng/mL). To explore

possible targets we tested several compounds for activity against a panel of *cis*- and *trans*-prenyl transferases (*cis*-FPPS, FPPS, DPPS, GGPPS, UPPS, CrtM and SQS) as well as against the menaquinone biosynthesis enzymes, MenA and MenG. Activity was seen against MenA and MenG and we proposed a structural model for the MenA active site, as well as a likely binding site for **1**. In addition, we found that menaquinone (MK-4) rescued both aerobic H37Rv *M. tuberculosis* cell growth, as well as the activity of non-replicating *M. tuberculosis* (streptomycin-starved 18b). We found that **1** as well as several analogs inhibited oxygen consumption in, *M. smegmatis* and there was a correlation between oxygen consumption and cell growth inhibition. We tested **1** and each of the 11 analogs for their effects on the PMF (ΔpH and Δψ) in fluorescence-based assays, as well as in some cases in intact cells (via ^{31}P NMR). The results obtained showed that **1** and the most potent cell growth inhibitors collapsed both Δψ and ΔpH and there were good correlations between experimental and predicted cell growth inhibition results based on MenA/ΔpH (*E.coli*) and respiration/Δψ *(M. smegmatis)*. Taken together, the results obtained suggested a model for **1**/analog activity in mycobacteria in which the increase in TMM levels seen on treatment with **1** have a contribution from (indirectly) inhibiting the TMM transporter MmpL3 by blocking the PMF/ATP biosynthesis. Overall, the results are of general interest since they indicate that **1** (and its analogs) can have diverse effects: on O_2-consumption/electron transport/MenA/MenG inhibition, Δψ, ΔpH, and ATP biosynthesis, likely helping to explain activity against non-MmpL3 containing pathogens such as *H. pylori*, *C. albicans* and here, *P. falciparum*. Moreover, the possibility of developing more potent compounds that can inhibit these targets is of general interest in the context of developing drug leads that are "resistance resistant", due to multi-targeting.

7.5 Materials and Methods

Chemical Syntheses: General methods.

All chemicals were reagent grade and were used as received. Moisture-sensitive reactions were performed under an inert atmosphere (dry nitrogen) with dried solvents. Reactions were monitored by TLC using Merck silica gel 60 F-254 thin-layer plates. Flash column chromatography was carried out on Merck silica gel 60 (230-400 mesh). ^{1}H

NMR and ^{13}C NMR spectra were recorded on Varian (Palo Alto, CA) Unity spectrometers at 400 and 500 MHz for ^{1}H and at 100 and 125 MHz for ^{13}C. Coupling constants (J) are reported in Hz. High-resolution mass spectra (HRMS) were recorded in the University of Illinois Mass Spectrometry Laboratory. Elemental analyses were carried out in the University of Illinois Microanalysis Laboratory. HPLC/MS was performed using an Agilent LC/MSD Trap XCT Plus system (Agilent Technologies, Santa Clara, CA) with a 1100 series HPLC system including a de-gasser, an auto-sampler, a binary pump, and a multiple-wavelength detector. All final compounds were ≥95% pure as determined by elemental analysis, analytical HPLC/MS analysis or qNMR analysis. qNMR spectra were recorded using Varian (Palo Alto, CA) 500 MHz Unity spectrometers with 1,3,5-trimethoxybenzene as the internal total-spin-count quantitation standard; 60 degree pulse excitation, 60 s recycle delay, 1.0 Hz line-broadening due to exponential multiplication, and 16 accumulations. qNMR data were processed using Mnova NMR software (Mestrelab, Escondido, CA). All NMR spectra (including qNMR spectra) are provided in the Supporting Information.

Enzyme Inhibition Assays.

MenA and MenG Inhibition. MenA and MenG inhibition assays were carried out using *M. smegmatis* membrane fragments[19]. Mycobacterial MenA assays were conducted as previously reported[19]. In addition, we used an expressed, purified *E. coli* MenA, as described below.

MenG Assay. Vitamins K1 and K2, and kanamycin were purchased from Sigma-Aldrich (St. Louis, MO). Authentic MK9 was purchased from Toronto Research Chemicals (TRC, Canada). S-adenosyl-L-[methyl-^{14}C] methionine (^{14}C-SAM) obtained from Perkin-Elmer (47 mCi/mmol). DMK8 was prepared from an *E. coli ΔubiE* mutant (CGSC #11636), which accumulates DMK8, and was purchased from the *E.coli* Genetic Stock Center, Yale University (http://cgsc.biology.yale. edu).

MenG assays were conducted using the membrane fractions prepared from *M. smegmatis* grown in 7H9 medium (supplemented with oleic acid, albumin, dextrose and 0.05% Tween 80). Washed cells were resuspended in Buffer A (50 mM MOPS pH 7.9, 5 mM $MgCl_2$, 5 mM DL-dithiothreitol (DTT), 10% glycerol (V/V)) and disrupted by probe sonication on ice with a Sanyo Soniprep 150 (10 cycles of 60 sec on and 90 sec off). The

whole cell lysate was centrifuged at 27,000 x g for 20 min at 4 °C. The supernatant was further centrifuged at 100, 000 x g (for 2 h at 4 °C) in an Optima TLX Ultracentrifuge (Beckman). The membrane-enriched pellet was washed with Buffer A followed by ultracentrifugation at 100, 000 g. The washed pellet was resuspended in Buffer A, divided into aliquots and frozen at -80 °C. The membrane protein concentration was estimated by using a BCA protein assay kit (Pierce).

Assay mixtures (100 µL) contained 100 mM Tris-HCL pH 8.0, 1 mM DTT, 5 mM $MgCl_2$, 0.1% CHAPS, 600 ng of DMK8, 40 µM radiolabelled SAM and varying concentrations of inhibitor **1** (0 to 25.0 µg/mL). Reactions were initiated by the addition of 50-100 µg of *M.smegmatis* membrane protein and incubated at 37 °C for 1 h. Reactions were stopped by the addition of 0.1 M acetic acid in methanol (0.5 mL) and radiolabelled products were extracted with hexane (2 x 3 mL). Pooled extracts were washed with 1 mL of water, evaporated to dryness under a N_2 stream and dissolved in $CHCl_3/CH_3OH$ (2:1 v/v). An aliquot was subjected to liquid scintillation counting (LS 6500, Beckman Coulter); an second aliquot and authentic standards (DMK8 and MK9) were subjected to reverse-phase TLC (Whatman KC 18F Silica gel 60 A) developed in acetone/water (97:3). Standards were visualized under UV light, distribution of radioactivity was detected by phosphorimaging (Typhoon TRIO, Amersham Biosciences) and quantified with ImageQuant TL v2005 software (Amersham Biosciences). IC_{50} values were calculated by using GraFit Software (Version 5.0.13).

Expression and purification of EcMenA. The gene encoding EcMenA with a N-terminal strep tag was amplified by polymerase chain reaction (PCR) with forward primer 5'-GACGACGACAAGATGAGCGCGTGGAGCCATCCGCAGTTTGAAAAAGGCGGTGGCAGCGCGGAGAATCTTTATTTTCAGGGCGCTGGTGC-3' and reverse primer 5'-GAGGAGAAG

CCCGGTTATTATGCTGCCCACTGGCTTAGGAATAT-3', and then cloned into the pET46 Ek/LIC vector. The recombinant plasmid was transformed to *E. coli* C43 (DE3) and the protein induced with 1 mM isopropyl-thiogalactopyranoside (IPTG) at 37 °C for 5 hours. The cell paste was harvested by centrifugation at 7,000 x g and re-

suspended in buffer A containing 25 mM Tris-HCl, pH 7.5, 150 mM NaCl and 20 mM imidazole. A cell lysate was prepared with a JNBIO® pressure cell (JN-3000 PLUS), the membrane and soluble proteins being separated by ultracentrifugation at 150,000 x g for 1.5 h. The resulting pellet was solubilized by incubation in buffer A supplemented with 1% (w/v) DDM detergent overnight at 4 °C. The latter solution was centrifuged (100,000 x g for 1 hr at 4 °C in a Beckman Ti70 rotor), and the supernatant loaded onto a Ni-NTA column and washed with buffer A containing 0.05% DDM. The buffer and gradient for the Ni-NTA column were 25 mM Tris, pH 7.5, 150 mM NaCl, 0.05% DDM and 20–500 mM imidazole. The protein was then loaded onto a Strep-Tactin® (IBA) column equilibrated with washing buffer containing 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, and 0.05% DDM and washed with 5 column volumes of washing buffer. EcMenA was finally eluted with eluting buffer containing 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.05% DDM, and 2.5 mM desthiobiotin. The purified protein was finally concentrated to 5 mg mL^{-1} in a 25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% DDM buffer.

EcMenA Inhibition Assay. Inhibition of EcMenA was carried out using an HPLC-based protocol. Typically, 2.5 μg EcMenA in 100 L reaction buffer (25 mM Tris-HCl, 0.1% Triton X-100, 250 μM $MgCl_2$, 10 mM DTT, pH=7.5) was incubated with inhibitors for 30 min at 22 ^oC. 1,4-dihydroxy-2-naphthoic acid (DHNA) and farnesyl diphosphate (FPP) were then added to the enzyme solution to a final concentration of 150 μM each. The reaction was incubated at 37 ^oC for 3 hrs before quenching with 50 μL 0.1 M acetic acid in methanol containing 50 μM menaquinone-4 (MK-4, Sigma-Aldrich) as an internal standard. The mixture was then extracted with 600 L hexane by vortexing. After centrifugation, 500 L organic layer was collected and dried under nitrogen then dissolved in 200 L methanol. 20 L of the methanol solution was then subjected to HPLC analysis (0.1 % formic acid in H_2O to 0.1 % formic acid in CH_3CN, UV: 325 nm, 250 μL/min). The amount of the MenA reaction product demethylmenaquinone-3 (DMMK-3) was determined by comparison of integrated peak areas between DMMK and the internal standard MK-4. IC_{50} values were estimated by using Origin 6.1 software to analyze the dose-response curves.

Cell lines. *Mycobacterium tuberculosis* ATCC® 27294™, *Mycobacterium smegmatis* ATCC® 700084™, *Bacillus subtilis* subsp. *subtilis* ATCC® 6051™, *E. coli* ATCC® 29425™, and *Saccharomyces cerevisiae* ATCC® 208352™ were purchased from the American Type Culture Collection. The *C. albicans* strain was CAI-4; the *P. falciparum* strain was 3D7 and the human cell line MCF-7 (breast adenocarcinoma), obtained from the National Cancer Institute.

***M. tuberculosis* Growth Inhibition Assay**. All 12 compounds (1, 3-13) were assayed for inhibition of *M. tuberculosis* cell growth as described previously[35].

Menaquinone Rescue Experiments With *M. tuberculosis* Treated With 1. We measured the activity of **1** against both actively growing *M. tuberculosis* (H37Rv) and non- replicating *M. tuberculosis* (streptomycin-starved 18b[25]) using a resazurin microplate reduction assay. The effects of menaquinone supplementation on the dose-response curves were investigated using medium that was supplemented with 0, 10, 100 and 1000 M menaquinone (MK-4, Sigma-Aldrich) in the presence of between 10 ng/mL and 10 μg/mL **1**. The activity of **1** against non-replicating 18b was determined after 7 days of drug exposure by plating the culture followed 28 days later by CFU counting after plating serial dilutions on 7H10 agar plates (Difco) .

***Candida albicans* Growth Inhibition Assay.** *C. albicans* growth inhibition was carried out according to a reported protocol[36] except that YPD media was used instead of RPMI 1640.

***E. coli* Growth Inhibition Assay.** IC_{50} values for *E. coli* growth inhibition were determined by using a broth micro-dilution method. An overnight culture of *E. coli* was diluted 50-fold into fresh Luria-Bertani (LB) broth and incubated to an OD_{600} of ~0.4. The culture was then diluted 500-fold into fresh LB medium and 100 L inoculated into each well of a 96 well flat bottom culture plate (Corning Inc., Corning, NY). The starting concentration of each compound was 0.3 mM and this was 2X serially diluted to 292 nM.

Plates were incubated for 3 h at 37 °C to mid-exponential phase. An MTT ((3-(4, 5-dimethylthiazole-2-yl)-2, 5-diphenylthtrazolium bromide) cell proliferation assay (ATCC) was then carried out to obtain bacterial viability dose-response curves. Briefly, 10 L MTT reagent was added into each well, followed by incubation for 2-4 hrs until a purple precipitate was visible. Then, 100 L detergent reagent was added and the plates incubated in the dark at 22 °C for 2 hours. Absorbance was measured at 570 nm and a non-linear regression analysis carried out using Origin 6.1 software

***B. subtilis* Growth Inhibition Assay.** IC_{50} values for *B. subtilis* growth inhibition were determined by using a micro-broth dilution method. A stationary starter culture of *B. subtilis* was diluted 50-fold into fresh LB broth and grown to an OD_{600} of ~0.4. The culture was then diluted 500-fold into fresh LB medium to give a working solution, then 100 L of working solution was transferred to each well of a 96-well flat bottom culture plate (Corning Inc., Corning, NY). Inhibitors were then added at 0.5 mM and 2x serial diluted to 500 nM the volume and solvent composition constant. Plates were incubated for 12-16 h at 37 °C and the absorbance at 600 nm determined. A non-linear regression analysis was carried out using Origin 6.1 in order to obtain the IC_{50} values.

***S. cerevisiae* Growth Inhibition Assay.**

The protocol was the same as for *B. subtilis* except that YPD instead of LB was used as the culture medium and the 96-well plates were incubated for 48 h instead of 12-16 h.

***Plasmodium falciparum* Growth Inhibition Assay.** We determined IC_{50} values for *P. falciparum* growth inhibition using the intra-erythrocytic assay described previously[37].

Human Cell Growth Inhibition Assay. The MCF-7 cell growth inhibition assay was carried out as described previously[38]. A broth micro-dilution method was used to determine the growth inhibition IC_{50} values. Compounds were half-log serial diluted using cell culture media into 96-well TC-treated round bottom plates (Corning Inc., Corning, NY). Cells were plated at a density of 5000 cells/well and then incubated under the same culture conditions for 2 days at which time an MTT ((3-(4,5-dimethylthi-azole-

2-yl)-2,5-diphenyltetrazolium bromide) cell proliferation assay (ATCC, Manassas, VA) was performed to obtain dose response curves.

Dehydrogenase Activities. Dehydrogenase activity in *M. smegmatis* membranes was measured by using the MTT ((3-(4,5-dimethylthiazole-2-yl)-2,5-diphenyltetrazolium bromide) reduction assay in the presence of 5 mM KCN. MTT reduction was followed at 570 nm, after addition of the different substrates (NADH, succinate, malate or lactate).

Oxygen Consumption. Oxygen concentration was monitored at 37 °C using a YSI model 53 oxygen electrode (Yellow Springs Instrument Co., Yellow Springs, OH) equipped with a temperature-controlled 1.8 mL electrode chamber. The reaction mixture consisted of sodium phosphate buffer, pH 7.5, 50 mM NaCl and 200-400 μg/mL membranes. The concentration of oxygen in the air-saturated buffer was taken to be 250 μM, and the reaction was initiated by injecting 200 μM NADH. The electron transport rates are expressed as mol NADH oxidized or mol O_2 (mol enzyme)$^{-1}$ s^{-1}. Membranes were incubated with different concentrations of inhibitors for 5 min prior to NADH addition.

Membrane Potential Measurements in Intact Cells. The effects of inhibitors on $\Delta\psi$ were determined by fluorescence quenching of the potential-sensitive probe 3,3'-dipropylthiodicarbocyanine (DisC3(5)). *M. smegmatis* were grown for 8 h in Middlebrook 7H9-ADC-Tween 80 medium and diluted to an OD_{600} of 0.3 in the same medium plus 10 mM glucose and 1 μM nigericin. Different concentrations of **1** and its analogs were added to the bacterial suspension and changes in fluorescence due to the disruption of $\Delta\psi$ were continuously monitored with a fluorescence spectrophotometer (FLUOstar Omega, BMG LABTECH) employing an excitation wavelength of 643 nm and an emission wavelength of 666 nm, at 30 ºC.

ATP/ADP determination. *M. smegmatis* were grown for 8 h in Middlebrook 7H9-ADC-Tween 80 and diluted to an OD_{600} of 2. Different concentrations of **1** and its analogs were added and ATP/ADP ratios determined (Abcam; ADP/ATP Ratio Assay

Kit, catalog number: ab65313) after 10 and 60 min of incubation at 37 ºC, 200 rpm. ATP and ADP were extracted from 50 μL cell suspension by adding trichloroacetic acid (TCA) to a final concentration of 0.5 %. After 5 min, TAE (Tris-acetic acid-EDTA) buffer was added to neutralize the system by diluting the sample 5-fold. The ATP and ADP cell concentrations were measured according to the manufacturer's protocol.

Inverted Membrane Vesicles (IMVs). *E. coli* IMVs were prepared by three passages through a pre-cooled French pressure cell at 20,000 psi. The lysate was centrifuged at 14,000 x g at 4 °C for 20 min to remove unbroken cells. The supernatant was centrifuged at 370,000 x g at 4 °C for 1 h and the pellet, consisting of the IMVs, was washed with 50 mM MOPS-KOH (pH 7.5), 2 mM $MgCl_2$. After the second centrifugation step, membranes were re-suspended in 50 mM MOPS-KOH (pH 7.5), 2 mM $MgCl_2$, 10 % glycerol and stored at -80°C.

Assay for ATP or Succinate-Driven Proton Translocation. Proton translocation into IMVs was measured by the decrease of ACMA fluorescence. The excitation and emission wavelengths were 410 nm and 480 nm, respectively. IMVs (0.1 mg/mL membrane protein) were pre-incubated at 37 °C in 10 mM HEPES-KOH (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$ containing 2 μM ACMA and the baseline monitored for five minutes. The reaction was then initiated by adding 1 mM ATP or 5 mM succinate. When the signal had stabilized, **1** or its analogs were added and proton translocation measured, fluorimetrically.

Determination of Δψ Collapse in IMVs. The Δψ-sensitive fluorophore Oxonol VI (1,5-bis(5-oxo-3-propylisoxazol-4-yl)pentamethine oxonol) was used to determine if **1** and its analogs were able to dissipate the membrane potential in IMVs. IMVs (0.1 mg/mL membrane protein) were added to assay buffer: 10 mM MOPS-KOH pH 7.5, 2 mM $MgCl_2$, 2 μM Oxonol VI. After a few seconds, 0.5 mM NADH was added to initiate respiration-dependent generation of Δψ (positive inside) and the resultant quenching of Oxonol VI fluorescence was monitored at 37 °C. The emission and excitation wavelengths were 599 nm and 634 nm, respectively. Uncoupling by inhibitors was

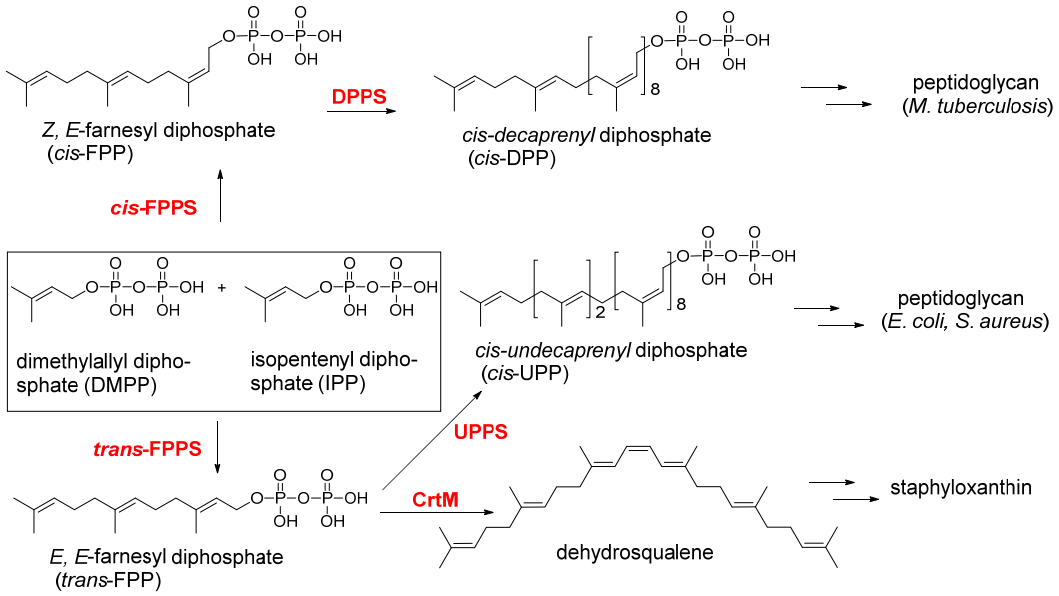
estimated based on their ability to dissipate the established $\Delta\psi$, measured as the de-quenching of the fluorescence signal.

Determination of pH by ^{31}P NMR Spectroscopy. *M. smegmatis* was grown to a cell density of 10^8 cells/mL in a total volume of 500 mL in a 4 L Erlenmeyer flask with constant shaking at 37 °C in Difco Middlebrook 7H9 media supplemented with oleic acid/albumin/dextrose and 0.05% Tween 80. Cells were harvested by centrifugation and the pellet washed twice with 5 mM phosphate buffer, pH 6.8. The cell pellet was then re-suspended in 200 L of the same buffer and 500 L of the resulting cell slurry transferred to a 5 mm NMR tube. Chemical shifts were reference with respect to 85% phosphoric acid in D_2O in a coaxial capillary. ^{31}P NMR spectra were obtained using a Varian INOVA 300 (at 121.5 MHz) using 60 degree pulse excitation, proton decoupling and a 1 s recycle time. 1024 scans were accumulated corresponding to ~ a 60 min total data acquisition time (without aeration). Spectra were analyzed as described elsewhere[39]. The peak corresponding to the -phosphate of ATP (at ~-10.5 ppm) and the inorganic phosphate peaks of interest (in the region of 0-1.5 ppm) were used to calculate the internal and external pH using the following equation, where d is the distance between the -phosphate of ATP and the inorganic phosphate peak, in ppm.

$$\mathrm{pH} = 6.75 + \log\frac{\mathrm{d}-10.85}{13.25-\mathrm{d}}$$

7.6 Schemes, Charts, Tables and Figures

Scheme 7.1 Illustration of several reactions of interest in isoprenoid biosynthesis in the systems investigated here. The enzymes in red were tested for inhibition by **1**. Cis-FPPS and trans-FPPS, UPPS and DPPS are not inhibited by **1** but CrtM is, and CrtM-**1** structures have been reported (PDB ID codes 4EA1, 4EA2) and serve as models for MenA inhibition.



Scheme 7.2 Menaquinone biosynthesis illustrated for MK-3. MK-8,9 are the abundant species in cells. MenA forms demethylmenaquinol (DMK) which spontaneously oxidizes to demethylmenaquinone. DMK is the substrate of MenG.

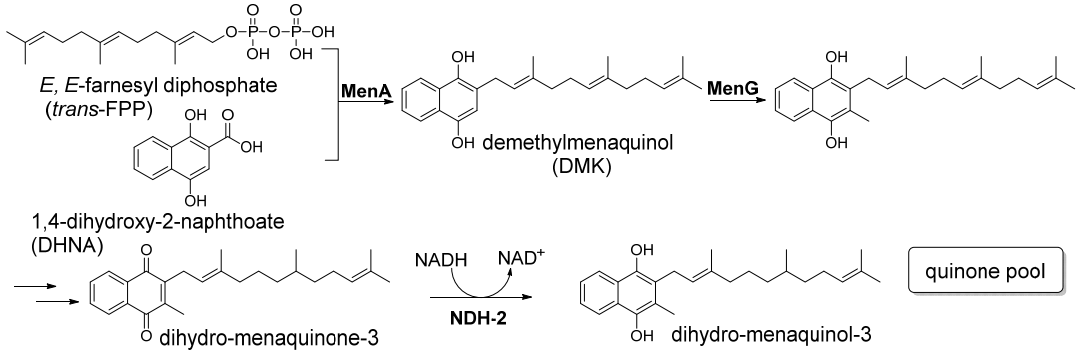


Chart 7.1 Structures of compounds investigated.

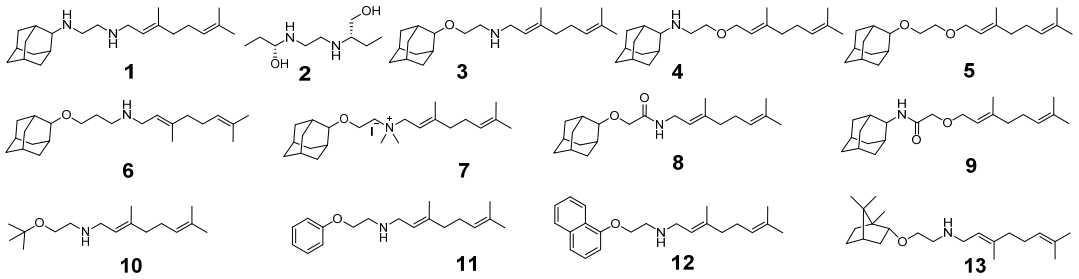


Table 7.1 Inhibition by 1 and analogs of *M. tuberculosis* (Mt), *M. smegmatis* (Ms), *S. aureus* (Sa), *B. subtilis* (Bs), *E. coli* (Ec), *S. cerevisiae* (Sc), *C. albicans* (Ca), *P. falciparum* (Pf) and human (MCF-7) cell growth.

	Mt[a]	Ms[a]	Sa[a]	Bs[b]	Ec[b]	Sc[b]	Ca[b]	Pf[b]	MCF-7[b]
1	0.1-0.2	3.1	>120	7.6	2.8	1.1	3.3	1	6
3	0.02-0.05	1.6	8	16	2.8	1.8	12	0.79	10
4	0.19	6.2	>120	>66	4.3	>33	>66	0.93	110
5	50	>50	>64	>66	>330	>330	>66	7.9	73
6	1.6	6.2	8	3	2.3	2.7	6.5	0.83	3.4
7	0.78	12	8	4.2	37	15	18	0.08	3.2
8	0.78	6.2	>120	>69	>350	>69	>69	6.2	12
9	12	25	>120	17	>350	>69	>69	3.8	9.3
10	6.2	50	>16	>50	>130	44	>50	3.2	45
11	6.2	12	32	17	60	3	20	14	6.3
12	6.2	12	8	2.4	15	3.7	12	>100	7.1
13	0.05	1.6	8	1.8	12	0.89	6.3	2	1.3

a. MIC_{90}, µg/mL; b. IC_{50}, µg/mL

Table 7.2 Enzyme, respiration and PMF (ΔpH, Δψ) inhibition results.

Entry	*Mycobacterium smegmatis*				*Escherichia coli*	
	MenA[a]	MenG[a]	respiration[b]	Δψ collapse[c]	MenA[d]	ΔpH collapse[e]
1	4.8	13	58	55	3.3	0.8
3	4	15	0.5	31	0.4	0.8
4	N.D.	N.D.	36	78	1.8	1.0
5	N.D.	N.D.	600	150	4.2	15
6	N.D.	N.D.	4.8	50	1.9	1.1
7	N.D.	N.D.	0.5	51	1.0	18
8	N.D.	N.D.	330	150	5.8	12
9	N.D.	N.D.	280	150	16	7.4
10	N.D.	N.D.	9500.0	150	4.6	4.7
11	N.D.	N.D.	2500.0	150	4.6	6.7
12	N.D.	N.D.	140.0	130	3.3	7.1
13	8	5.7	0.3	44	3.2	0.8

a. IC_{50} in μM, *M. smegmatis* membrane fraction (Figure 7.3); b. IC_{50} in μM, from methylene blue reduction assay (Figure 7.7); c. IC_{50} in μM, from DisC3(5) assay (Figure 7.12); d. IC_{50} in μM, expressed *E. coli* MenA (Figure 7.4); e. IC_{50} in μM, measured with *E. coli* IMVs (Figures 7.8 and 7.9). N.D.: not determined

Table 7.3 SCORECON analysis of MtMenA. Aspartate-rich motifs are highlighted in bold.

Ranking	SCORECONS Scores	Residues No.	Residue	Ranking	SCORECONS Scores	Residues No.	Residue
1	**1.000**	**196**	**D**	14	0.785	196	G
2	0.976	142	G	15	0.769	142	G
3	0.966	209	G	16	0.765	209	L
4	0.923	151	G	17	0.764	151	N
5	0.892	138	Y	18	0.757	138	E
6	0.832	201	K	19	0.756	201	Y
7	**0.823**	**200**	**D**	**20**	0.751	200	F
8	**0.820**	**192**	**D**	**21**	**0.748**	**192**	**D**
9	0.817	206	V	22	0.737	206	N
10	0.802	188	N	23	**0.737**	**188**	**D**
11	0.799	192	T	24	0.715	192	G
12	0.794	206	F	25	**0.672**	**206**	**D**
13	0.787	188	R	26	0.656	188	L

Figure 7.1 Schematic illustrations of series (in the same metabolic pathway), parallel (unrelated pathways or DNA/membrane targets) or network (series and parallel target) inhibition.

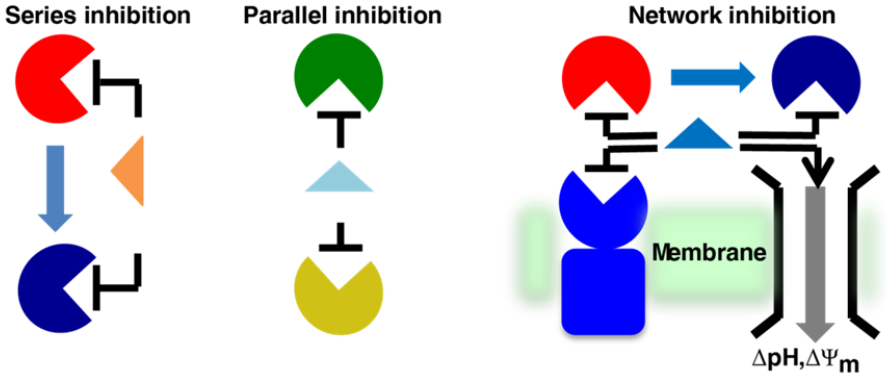


Figure 7.2 Inhibition of cell growth for *M. tuberculosis* (Mt), *M. smegmatis* (Ms), *S. aureus* (Sa), *B. subtilis* (Bs), *E. coli* (Ec), *S. cerevisiae* (Sc), *C. albicans* (Ca), *P. falciparum* (Pf) and a human cell line (MCF-7) by **1**, **3**-**13**. (A) Heat map. Red=strong inhibition; yellow=moderate inhibition; green=weak/no inhibition; (B) Correlation R values between cell growth inhibition pIC_{50} ($-\log_{10}IC_{50}$) between all systems investigated.

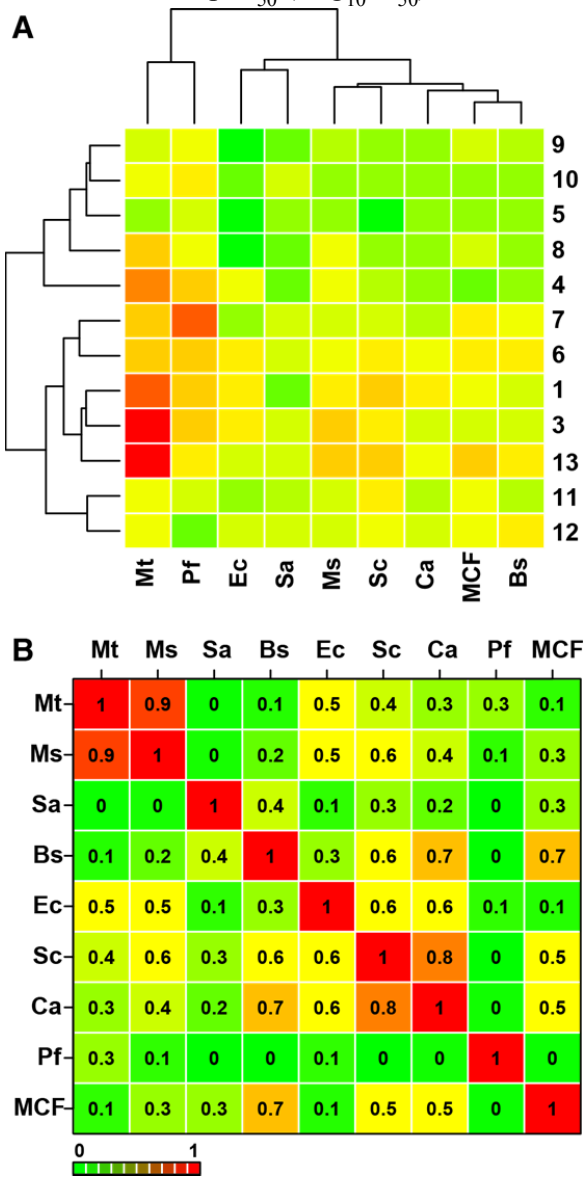


Figure 7.3 (A) MenA inhibition using *M. smegmatis* membrane fragments. 3 (blue, IC_{50} = 4.0 µM), **1** (black, IC_{50} = 4.8 µM), 1**3** (red, IC_{50} = 8.0 µM). (B) Raw data for MenG inhibition by **1**, **3** and **13** used to generate data in (C). The assay requires the addition of exogenous demethylmenaquinone, which was obtained from an *E. coli* MenG knockout (KO). Inhibitors were added at the indicated concentrations. (C) MenG inhibition results. **3** (blue, IC_{50} = 14.7 µM), **1** (black, IC_{50} = 12.6 µM), **13** (red, IC_{50} = 5.7 µM).

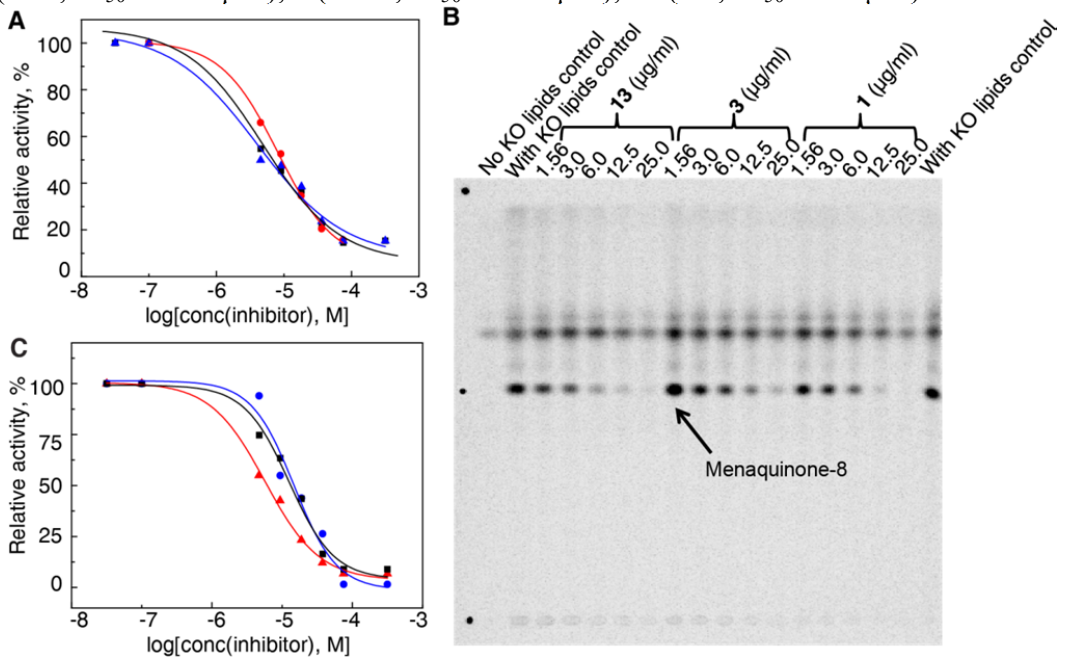


Figure 7.4. *E.coli* MenA inhibition by 1 and its analogs.

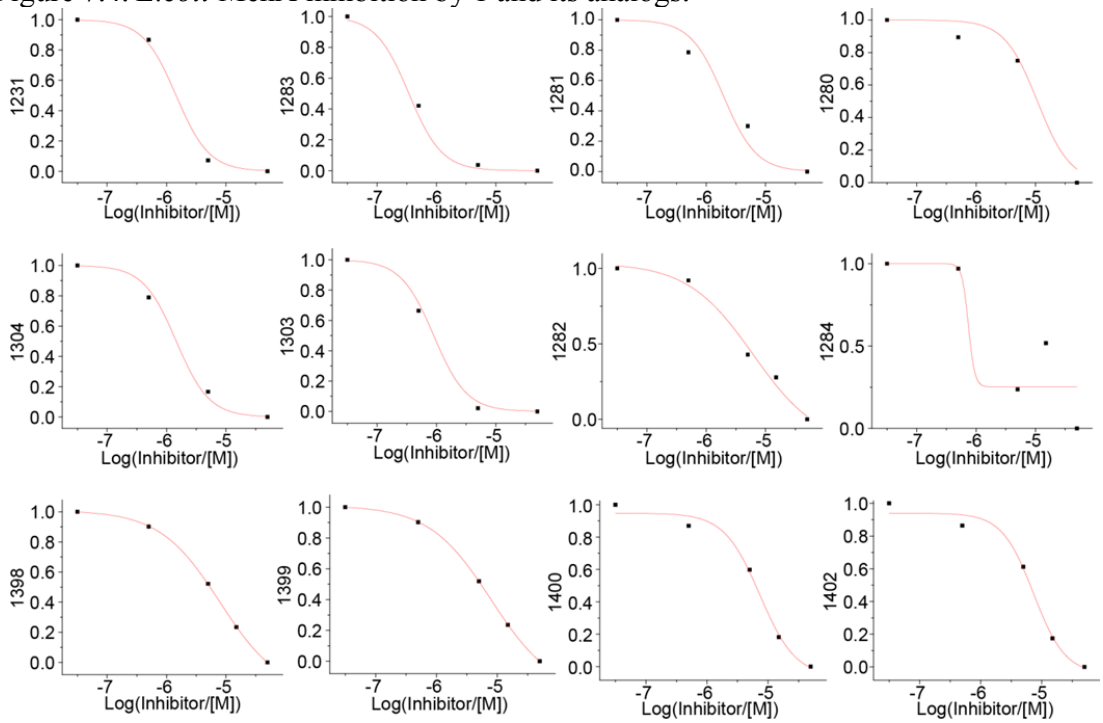


Figure 7.5 (A) Transmembrane helices predictions for MtMenA. (B) Transmembrane helices in Phyre2 model of MtMenA (helices S1, S8 and S9 from (A) are not modeled). Orange: Asp rich motifs. (C) MenA model (cyan) and CrtM (green, PDB: 4EA1, N and C terminal helices are removed). Blue: Asp rich motifs in CrtM. CrtM structure contains SQ109 (2 conformers), shown as magenta spheres.

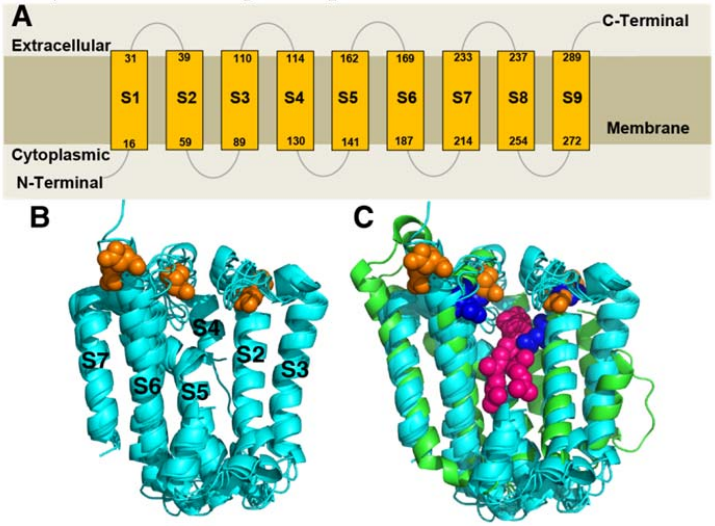


Figure 7.6 Menaquinone rescue experiments. (A) Aerobic *M. tuberculosis* H37Rv growth inhibition in the presence of increasing MK-4 concentrations, measured by REMA. (B) as (A), but with non-replicating *M. tuberculosis* 18b. (C) *M. tuberculosis* 18b cells were plated after 7 days of drug exposure with (grey) or without (black) MK-4. Colony Forming Unit counts were assessed after one month of incubation. Concentrations are in µg/ml.

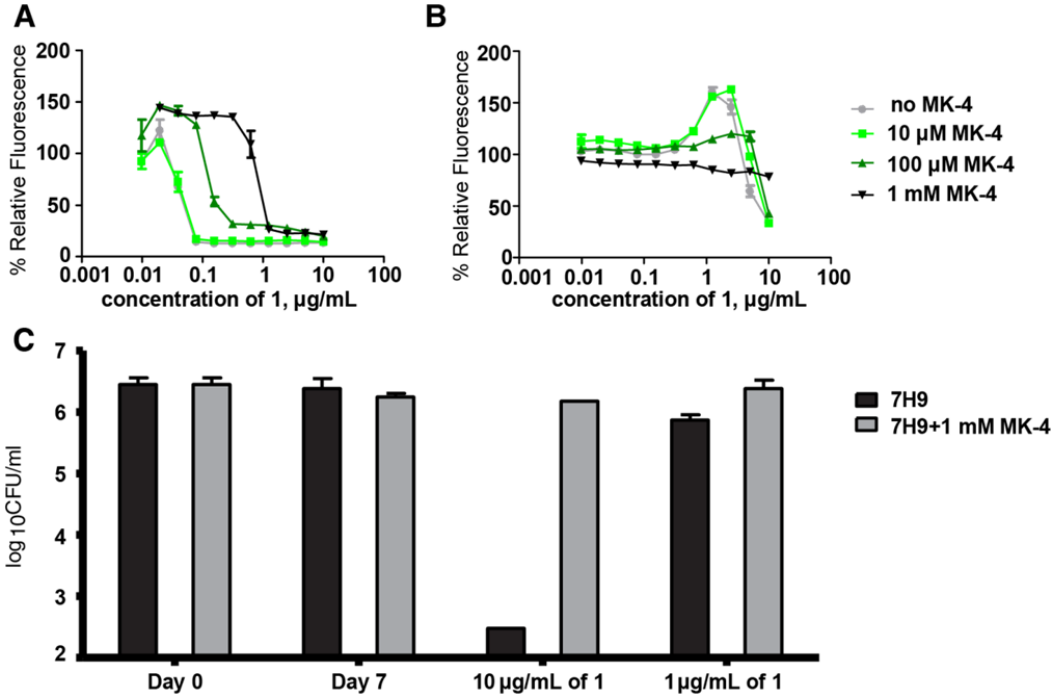


Figure 7.7 Rates of oxygen depletion by *M. smegmatis* cells in the presence or absence of 25 μg/mL compounds, measured by using the methylene blue reduction assay[1] monitoring at 670 nm. Data are expressed as average ± standard deviation of three independent experiments.

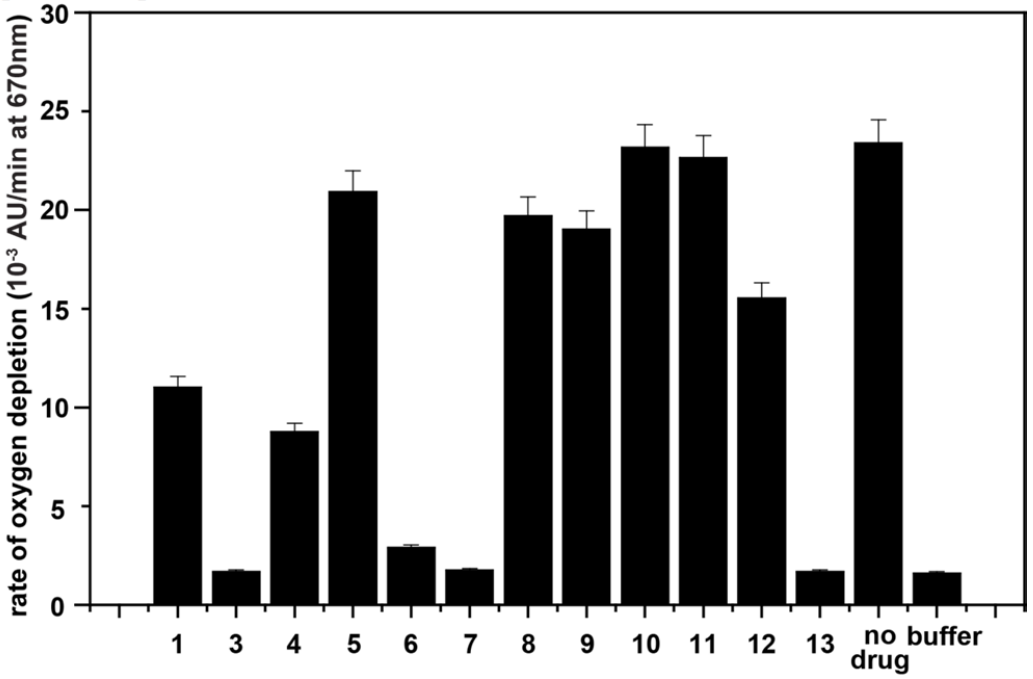


Figure 7.8 De-energization of *E. coli* inverted membrane vesicles (IMVs) by 1 and its analogs. Quenching of ACMA fluorescence in IMV was observed after the addition of ATP, which results in a decrease of fluorescence intensity (top left panel). Addition of inhibitors results in an increase of fluorescence intensity (I1). Data are representative of results of three separate experiments.

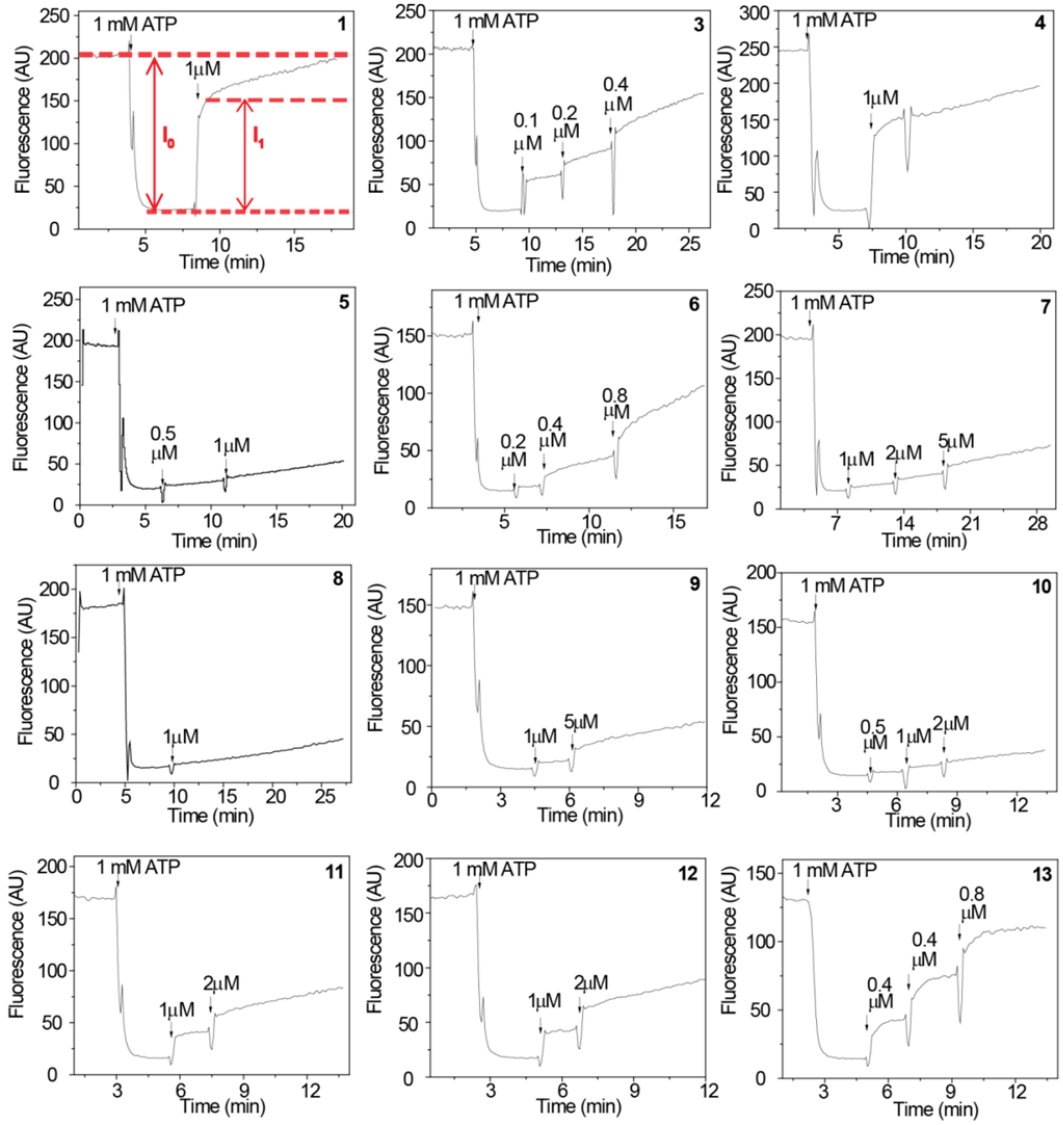


Figure 7.9 Increase of ACMA fluorescence in *E. coli* IMV in the presence or absence of 1 μM compounds, expressed as an I_0/I_1 percentage. Data are expressed as average ± standard deviation of five independent experiments.

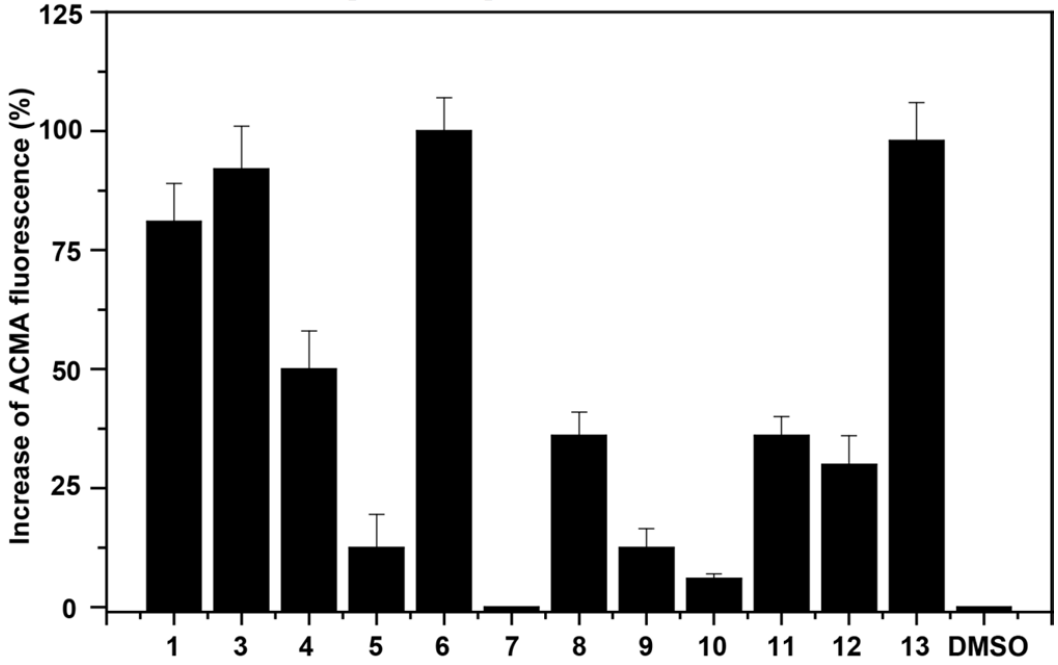


Figure 7.10 (A) ΔpH collapse in *E. coli* IMVs by a known uncoupler SF6847. (B) ΔpH collapse in *E. coli* IMVs by **1** and **5**. (C) Effects of **1** and analogs on Δψ in *M. smegmatis* cells. (D) Effects of **3** on ATP biosynthesis of effect in *M. smegmatis* cells.

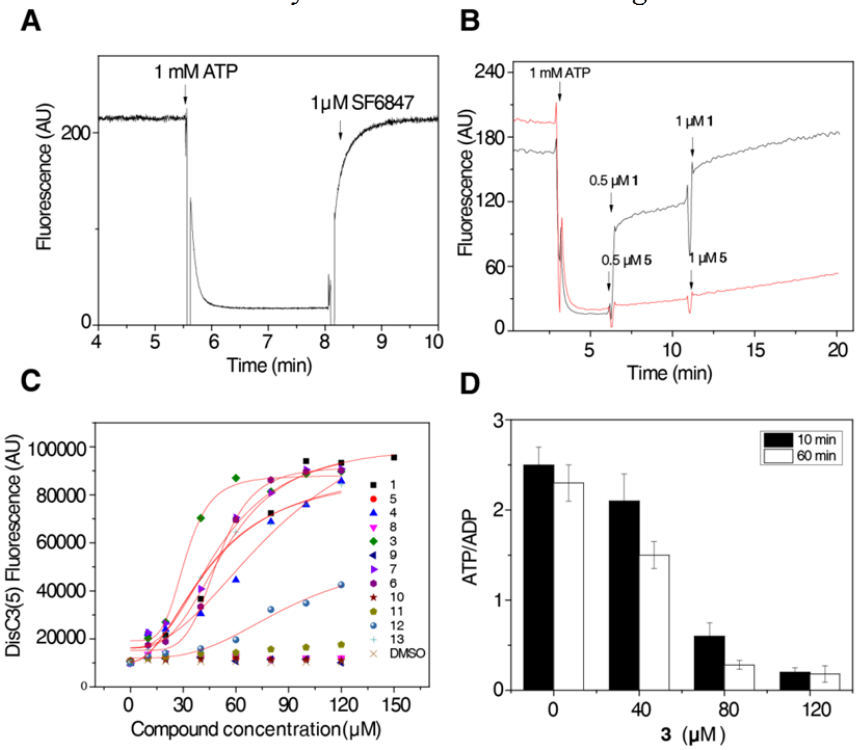


Figure 7.11 (A) Increase of Oxonol VI fluorescence in *E. coli* IMVs in the presence of 1 and its analogs. Oxonol VI fluorescence in IMVs was quenched after the addition of NADH. Upon the addition of 1 μM compounds, the Oxonol VI fluorescence increased due to a dissipation of the membrane potential. Data are expressed as average ± standard deviation of three independent experiments. (B) Correlation between increase of ACMA fluorescence and Oxonol fluorescence.

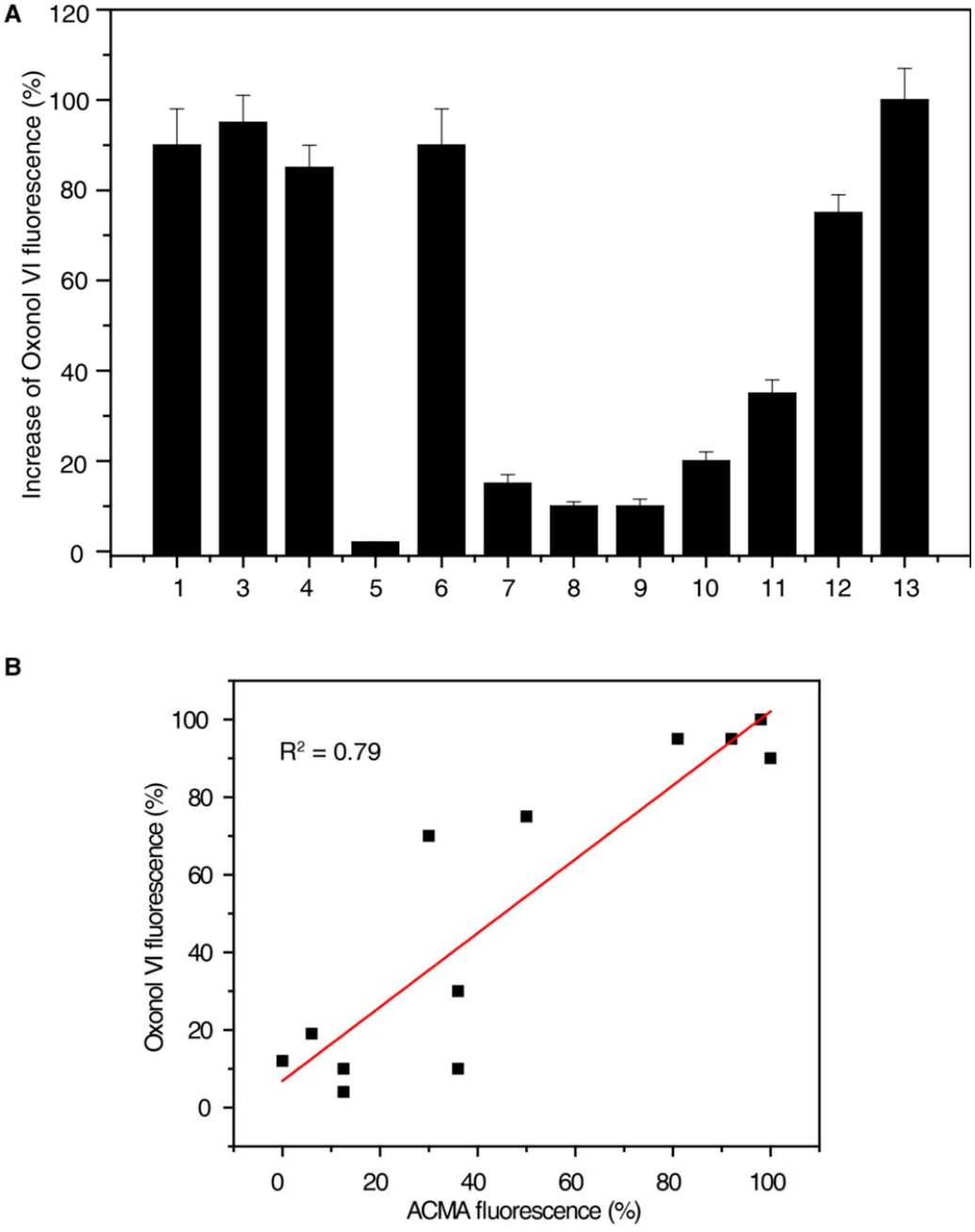


Figure 7.12 Δψ dissipation in *M. smegmatis* cells. (A) DisC3(5) fluorescence was quenched in the presence of cells and after the addition of 50 μM of each compound, the fluorescence increased due to the dissipation of the PMF. Data are expressed as average ± standard deviation of three independent experiments. (B) Effect of **1** concentration on Δψ dissipation in *M. smegmatis* cells.

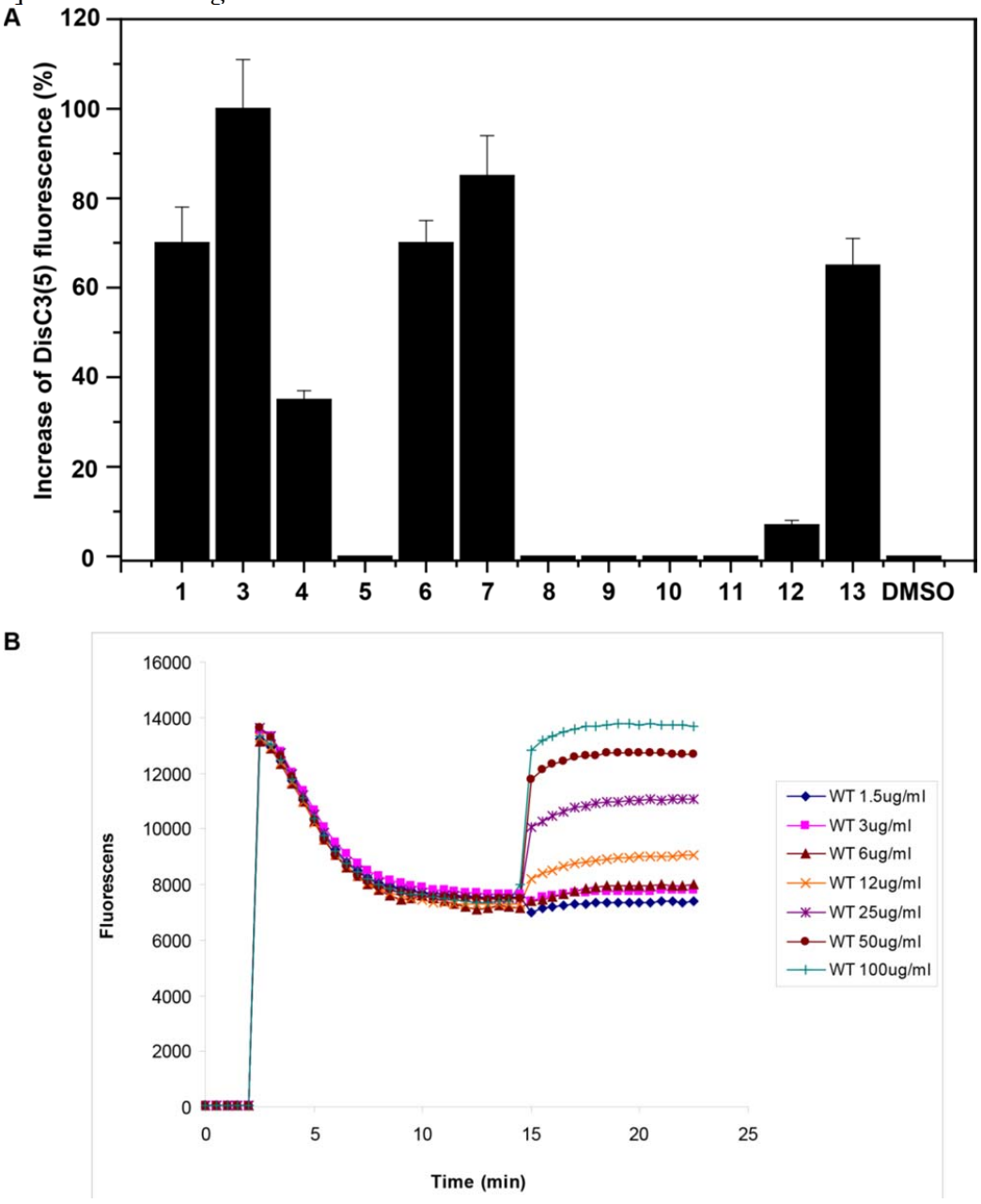
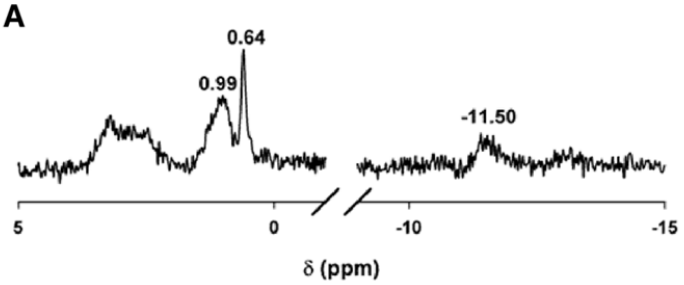


Figure 7.13 ^{31}P NMR-derived ΔpH collapse results in *M. smegmatis*. (A) ^{31}P NMR spectrum, wt *M. smegmatis*. (B) ΔpH results with various inhibitors, ionophores. **1** and **13** collapse the ΔpH.

A



B

Drug	pH outside	pH inside	ΔpH
M. smegmatis (No drug)	6.82	7.09	0.26
CCCP (5 μM)	6.85	6.85	0
Nigericin (5 μM)	6.82	6.82	0
1 (50 μg/mL)	6.81	6.81	0
13 (10 μg/mL)	6.83	6.83	0
Valinomycin (33 μM)	6.83	7.07	0.24

Figure 7.14 Experimental (red circles) and computed (colored plane) results for cell growth inhibition based on Equation 1. (A) *E. coli* cell growth inhibition predicted using MenA, ΔpH collapse (IC_{50}s in μM, R^2 for the model=0.77). (B) *M. smegmatis* growth inhibition using respiration (methylene blue assay) and Δψ collapse (MIC_{90} and IC_{50}s in μM, R^2 for the model=0.64).

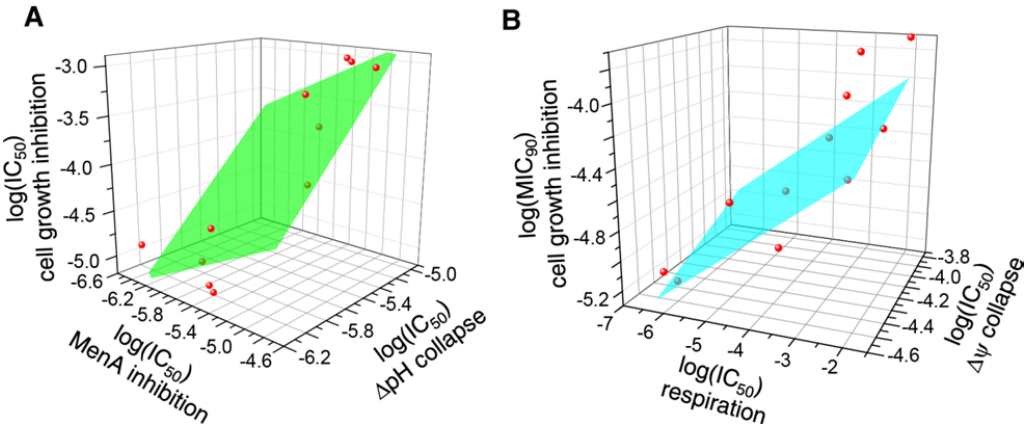


Figure 7.15 Molecular models for MmpL3, MmpL11. (A) Phyre2 structure predictions for MmpL3. (B) Phyre2 structure predictions for MmpL11. (C) Phyre2 predictions showing hydrophobic residues (white/grey) and their proposed relation to the membrane for MmpL3. (D) Phyre2 predictions showing hydrophobic residues (white/grey) and their proposed relation to the membrane for MmpL11. (E) Structure of representative *M. tuberculosis* growth inhibitors that are thought to target MmpL3 and (F) sites of resistance mutations (blue spheres) in MmpL3.

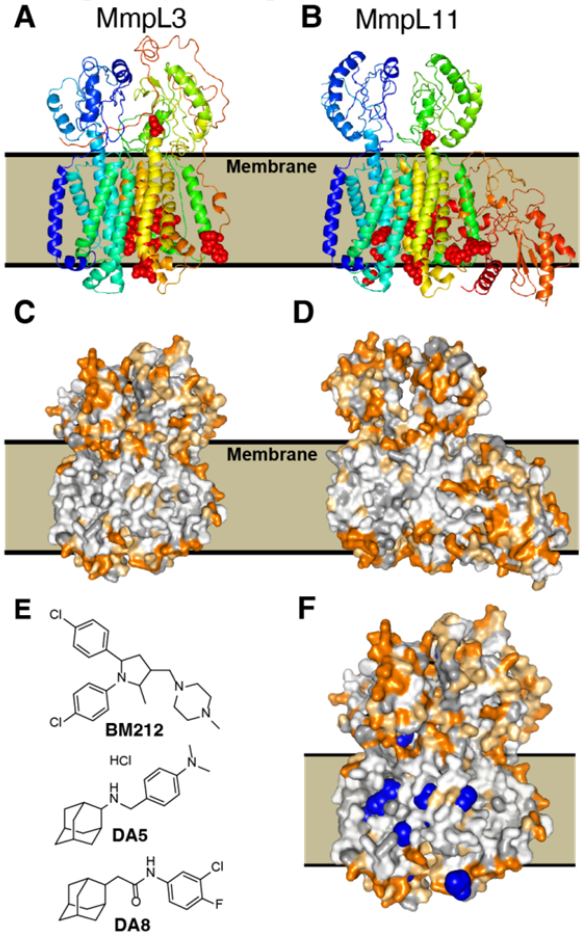
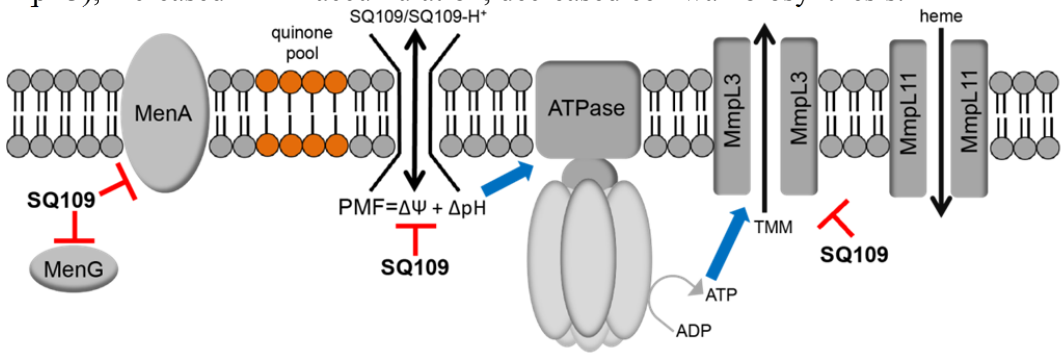


Figure 7.16 Schematic illustration of proposed sites of action of SQ109 and its analogs. MenA, MenG targeting can affect respiration/electron transfer; PMF (ΔpH, Δψ) collapse leads to decreased ATP biosynthesis, reduction in PMF/ATP-powered transporters (e.g. MmpL3), increased TMM accumulation, decreased cell wall biosynthesis.



7.7 References

1. US Centers for Disease Control (2013) Antibiotic Resistance Threats in the United States, 2013. http://www.cdc.gov/features/AntibioticResistanceThreats/index.htlm
2. World Health Organization. http://www.who.int/tb/publications/global_report/en/index.html
3. National Research Council. *Treating Infectious Diseases in a Microbial World: Report of Two Workshops on Novel Antimicrobial Therapeutics*. Washington, DC: The National Academies Press, 2006
4. Gray, K. C.; Palacios, D. S.; Dailey, I.; Endo, M. M.; Uno, B. E.; Wilcock, B. C.; Burke, M. D. Amphotericin primarily kills yeast by simply binding ergosterol. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** 109, 2234-2239.
5. Fischbach, M. A. Combination therapies for combating antimicrobial resistance. *Curr. Opin. Microbiol.* **2011,** 14, 519-523.
6. Silver, L. L. Multi-targeting by monotherapeutic antibacterials *Nat. Rev. Drug Discov.* **2007,** 6, 126-126.
7. Morphy, J. R. The Challenges of Multi-Target Lead Optimization. In *Designing Multi-Target Drugs*, 2012; pp 141-154.
8. Sacksteder, K. A.; Protopopova, M.; Barry, C. E.; Andries, K.; Nacy, C. A. Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action. *Future Microbiol.* **2012,** 7, 823-837.
9. Martin, M. B.; Arnold, W.; Heath, H. T.; Urbina, J. A.; Oldfield, E. Nitrogen-containing bisphosphonates as carbocation transition state analogs for isoprenoid biosynthesis. *Biochem. Bioph. Res. Commun.* **1999,** 263, 754-758.
10. Protopopova, M.; Hanrahan, C.; Nikonenko, B.; Samala, R.; Chen, P.; Gearhart, J.; Einck, L.; Nacy, C. A. Identification of a new antitubercular drug candidate, SQ109, from a combinatorial library of 1,2-ethylenediamines. *J. Antimicrob. Chemother.* **2005,** 56, 968-974.
11. Tahlan, K.; Wilson, R.; Kastrinsky, D. B.; Arora, K.; Nair, V.; Fischer, E.; Barnes, S. W.; Walker, J. R.; Alland, D.; Barry, C. E.; Boshoff, H. I. SQ109 Targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemther.* **2012,** 56, 1797-1809.
12. La Rosa, V.; Poce, G.; Canseco, J. O.; Buroni, S.; Pasca, M. R.; Biava, M.; Raju, R. M.; Porretta, G. C.; Alfonso, S.; Battilocchio, C.; Javid, B.; Sorrentino, F.; Ioerger, T. R.; Sacchettini, J. C.; Manetti, F.; Botta, M.; De Logu, A.; Rubin, E. J.; De Rossi, E. MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. *Antimicrob. Agents Chemther.* **2012,** 56, 324-331.
13. Makobongo, M. O.; Einck, L.; Peek, R. M.; Merrell, D. S. In vitro characterization of the anti-bacterial activity of SQ109 against *Helicobacter pylori*. *Plos One* **2013,** 8, e68917.
14. Barbosa, e. a. In vitro antifungal susceptibility testing of drug candidate SQ109 *Candida albicans*. In *Intersicence Conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, San Francisco, CA, 2006.
15. Oldfield, E.; Lin, F. Y. Terpene biosynthesis: modularity rules. *Angew. Chem. Int. Edit.* **2012,** 51, 1124-1137.

16. Meganathan, R. Biosynthesis of menaquinone (vitamin K-2) and ubiquinone (coenzyme Q): A perspective on enzymatic mechanisms. *Vitam. Horm.* **2001,** 61, 173-218.
17. Kurosu, M.; Narayanasamy, P.; Biswas, K.; Dhiman, R.; Crick, D. C. Discovery of 1,4-dihydroxy-2-naphthoate prenyltransferase inhibitors: new drug leads for multidrug-resistant gram-positive pathogens *J. Med. Chem.* **2007,** 50, 5048-5048.
18. Kurosu, M.; Crick, D. C. MenA is a promising drug target for developing novel lead molecules to combat *Mycobacterium tuberculosis*. *Med. Chem.* **2009,** 5, 197-207.
19. Dhiman, R. K.; Mahapatra, S.; Slayden, R. A.; Boyne, M. E.; Lenaerts, A.; Hinshaw, J. C.; Angala, S. K.; Chatterjee, D.; Biswas, K.; Narayanasamy, P.; Kurosu, M.; Crick, D. C. Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence. *Mol. Microbiol.* **2009,** 72, 85-97.
20. Zhang, Y. H.; Cao, R.; Yin, F.; Hudock, M. P.; Guo, R. T.; Krysiak, K.; Mukherjee, S.; Gao, Y. G.; Robinson, H.; Song, Y.; No, J. H.; Bergan, K.; Leon, A.; Cass, L.; Goddard, A.; Chang, T. K.; Lin, F. Y.; Van Beek, E.; Papapoulos, S.; Wang, A. H. J.; Kubo, T.; Ochi, M.; Mukkamala, D.; Oldfield, E. Lipophilic bisphosphonates as dual farnesyl/geranylgeranyl diphosphate synthase inhibitors: an X-ray and NMR investigation. *J. Am. Chem. Soc.* **2009,** 131, 5153-5162.
21. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001,** 305, 567-580.
22. Kelley, L. A.; Sternberg, M. J. E. Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **2009,** 4, 363-371.
23. Lin, F. Y.; Liu, Y. L.; Li, K.; Cao, R.; Zhu, W.; Axelson, J.; Pang, R.; Oldfield, E. Head-to-head prenyl tranferases: anti-infective drug targets. *J. Med. Chem.* **2012,** 55, 4367-72.
24. Valdar, W. S. Scoring residue conservation. *Proteins* **2002,** 48, 227-41.
25. Sala, C.; Dhar, N.; Hartkoorn, R. C.; Zhang, M.; Ha, Y. H.; Schneider, P.; Cole, S. T. Simple model for testing drugs against nonreplicating *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemther.* **2010,** 54, 4150-4158.
26. Felser, A.; Blum, K.; Lindinger, P. W.; Bouitbir, J.; Krahenbuhl, S. Mechanisms of hepatocellular toxicity associated with dronedarone-a comparison to amiodarone. *Toxicol. Sci.* **2013,** 131, 480-490.
27. Garlid, K. D.; Nakashima, R. A. Studies on the mechanism of uncoupling by amine local-anesthetics - evidence for mitochondrial proton transport mediated by lipophilic ion-pairs. *J. Biol. Chem.* **1983,** 258, 7974-7980.
28. Moreno-Sanchez, R.; Bravo, C.; Vasquez, C.; Ayala, G.; Silveira, L. H.; Martinez-Lavin, M. Inhibition and uncoupling of oxidative phosphorylation by nonsteroidal anti-inflammatory drugs - Study in mitochondria, submitochondrial particles, cells, and whole heart. *Biochem. Pharmacol.* **1999,** 57, 743-752.
29. Yu, E. W.; McDermott, G.; Zgurskaya, H. I.; Nikaido, H.; Koshland, D. E. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **2003,** 300, 976-980.

30. Haagsma, A. C.; Podasca, I.; Koul, A.; Andries, K.; Guillemont, J.; Lill, H.; Bald, D. Probing the interaction of the diarylquinoline TMC207 with its target mycobacterial ATP synthase. *Plos One* **2011,** 6, e23575.
31. Moon, R. B.; Richards, J. H. Determination of intracellular ph by P-31 magnetic-resonance. *J. Biol. Chem.* **1973,** 248, 7276-7278.
32. Mukkamala, D.; No, J. H.; Cass, L. A.; Chang, T. K.; Oldfield, E. Bisphosphonate inhibition of a Plasmodium farnesyl diphosphate synthase and a general method for predicting cell-based activity from enzyme data. *J. Med. Chem.* **2008,** 51, 7827-7833.
33. Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L. Evolutionary relationships of microbial aromatic prenyltransferases. *Plos One* **2011,** 6, e27336.
34. Meng, J.; Wang, F.; Zheng, Y.; Peng, X.; Zhou, H.; Xiao, X. An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J.* **2009,** 3, 106-16.
35. Rao, S. P.; Lakshminarayana, S. B.; Kondreddi, R. R.; Herve, M.; Camacho, L. R.; Bifani, P.; Kalapala, S. K.; Jiricek, J.; Ma, N. L.; Tan, B. H.; Ng, S. H.; Nanjundappa, M.; Ravindran, S.; Seah, P. G.; Thayalan, P.; Lim, S. H.; Lee, B. H.; Goh, A.; Barnes, W. S.; Chen, Z.; Gagaring, K.; Chatterjee, A. K.; Pethe, K.; Kuhen, K.; Walker, J.; Feng, G.; Babu, S.; Zhang, L.; Blasco, F.; Beer, D.; Weaver, M.; Dartois, V.; Glynne, R.; Dick, T.; Smith, P. W.; Diagana, T. T.; Manjunatha, U. H. Indolcarboxamide is a preclinical candidate for treating multidrug-resistant tuberculosis. *Sci. Transl. Med.* **2013,** 5, 214ra168.
36. Grzegorzewicz, A. E.; Pham, H.; Gundi, V. A.; Scherman, M. S.; North, E. J.; Hess, T.; Jones, V.; Gruppo, V.; Born, S. E.; Kordulakova, J.; Chavadi, S. S.; Morisseau, C.; Lenaerts, A. J.; Lee, R. E.; McNeil, M. R.; Jackson, M. Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. *Nat. Chem. Biol.* **2012,** 8, 334-41.
37. Owens, C. P.; Chim, N.; Goulding, C. W. Insights on how the *Mycobacterium tuberculosis* heme uptake pathway can be used as a drug target. *Future Med. Chem.* **2013,** 5, 1391-403.
38. Owens, C. P.; Chim, N.; Graves, A. B.; Harmston, C. A.; Iniguez, A.; Contreras, H.; Liptak, M. D.; Goulding, C. W. The Mycobacterium tuberculosis secreted protein Rv0203 transfers heme to membrane proteins MmpL3 and MmpL11. *J. Biol. Chem.* **2013,** 288, 21714-28.
39. Gruppo, V.; Johnson, C. M.; Marietta, K. S.; Scherman, H.; Zink, E. E.; Crick, D. C.; Adams, L. B.; Orme, I. M.; Lenaerts, A. J. Rapid microbiologic and pharmacologic evaluation of experimental compounds against Mycobacterium tuberculosis. *Antimicrob. Agents Chemther.* **2006,** 50, 1245-1250.
40. Reference Method for Broth dilution Antifungal Susceptibility Testing of Yeasts. In *Approved Standard-Third Edition*, Clinical and Laboratory Standards Institute: 2008.
41. No, J. H.; de Macedo Dossin, F.; Zhang, Y.; Liu, Y. L.; Zhu, W.; Feng, X.; Yoo, J. A.; Lee, E.; Wang, K.; Hui, R.; Freitas-Junior, L. H.; Oldfield, E. Lipophilic analogs of zoledronate and risedronate inhibit Plasmodium geranylgeranyl diphosphate synthase (GGPPS) and exhibit potent antimalarial activity. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** 109, 4058-4063.

42. Compain, J. D.; Mialane, P.; Marrot, J.; Secheresse, F.; Zhu, W.; Oldfield, E.; Dolbecq, A. Tetra- to dodecanuclear oxomolybdate complexes with functionalized bisphosphonate. *Chem. Eur. J.* **2010,** 16, 13741-13748.
43. Navon, G.; Ogawa, S.; Shulman, R. G.; Tamane, T. High-resolution P-31 nuclear magnetic-resonance studies of metabolism in aerobic escherichia-coli-cells. *Proc. Natl. Acad. Sci. U.S.A.* **1977,** 74, 888-891.